Examining Theories of Ventromedial Prefrontal Cortex Function

by

Elizabeth Zabriskie Wheeler

B.A., Wellesley College, 1997

M.S., University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Elizabeth Z. Wheeler

It was defended on

April 20th, 2006

and approved by

Lesley Fellows, M.D., D.Phil.

Julie Fiez, Ph.D.

William Klein, Ph.D.

Greg Siegle, Ph.D.

Walter Schneider, Ph.D.
Dissertation Director

EXAMINING THEORIES OF VENTROMEDIAL PREFRONTAL CORTEX FUNCTION

Elizabeth Z. Wheeler, PhD

University of Pittsburgh, 2006

The ventromedial prefrontal cortex (VMPFC) is an intriguing brain region which sends output to and receives input from memory, emotion and reward related structures such as the amygdala, hippocampus, and caudate nucleus. Humans with lesions to the VMPFC on the surface seem to function normally and most have normal intelligence. However, in high-level tasks blending affect and decision-making, they are often highly impaired. This thesis concerns three behavioral experiments of patients with VMPFC damage which contrast and examine hypotheses of VMPFC function. In Experiment 1, the hypothesis that the VMPFC is involved in representing social knowledge was tested with more rigorous methods and a non social control task. Results did not support a specific role of the VMPFC as being uniquely involved in social knowledge. In Experiments 2 & 3, the hypothesis that VMPFC is involved in rapid reversal of stimulus-reinforcer associations was examined in detail. A gambling task and a probabilistic learning task helped discriminate punishment versus reward processing. Experiment 2 revealed normal performance of VMPFC patients in a rewards-only reversal task, in contrast to performance on previous gambling tasks with both reversal and punishment. Experiment 3 added to this evidence for a special function in punishment processing by examining learning from punishment versus learning from reward. Results revealed deficits in punishment learning, but not reward learning, after damage to the VMPFC. In conclusion, these experiments suggest a special role for the VMPFC in punishment processing, especially when a change in stimulus choice is indicated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

The author would like to thank the members of the dissertation committee for their helpful discussions and guidance. The author is also indebted to Dr. Frank for use of and guidance on modifications to his probabilistic learning task. This work would also have been greatly stymied without the help of the Farah Laboratory, Fellows Laboratory, Schneider Laboratory, and last, but definitely not least, the author's husband and extended family.

I. INTRODUCTION

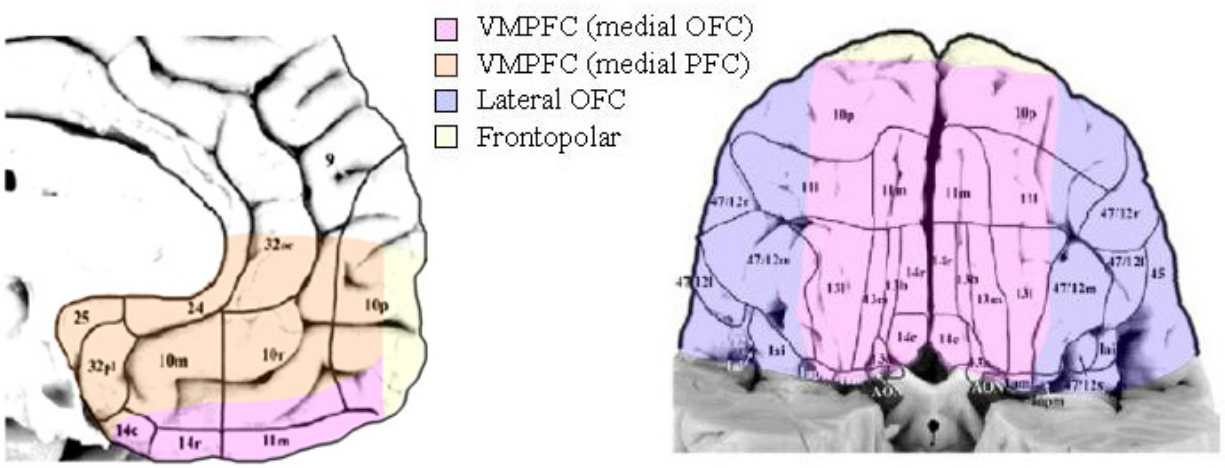
The ventromedial prefrontal cortex (VMPFC; see Figure 1) is an area of the brain which lies at the base of the frontal cortex behind the bridge of the nose. The VMPFC is important as a nexus of emotion- and reward-related structures, such as the amygdala, nucleus accumbens and caudate nucleus, with memory-related structures, such as the ento-rhinal cortex and hippocampus (Ongur & Price, 2000). The VMPFC is closely associated with a variety of sensory modalities (gustatory, olfactory, visual, visceral). However, it can also be thought of as a subset of the association areas of the prefrontal region, known to be involved in higher-level or executive functioning (Ongur & Price, 2000). Damage to the VMPFC can result in subtle behavioral changes. On the surface, these patients often seem normal, but may be unable to keep a job or function independently. In executive tasks combining affect and decision-making, they are often impaired. However, the origin of these deficits is as yet unclear. Several hypotheses have been proposed for the function of VMPFC, including such diverse ideas as social knowledge storage, affective working memory, and flexible reinforcer processing. The following experiments will attempt to examine some of the existing theories in order to discover whether the evidence is weighted towards one in exclusion of the others.

This thesis starts with an introduction to the VMPFC and related areas, as well as subsets of the VMPFC. A brief introduction to three theories of VMPFC function will follow. Then, the three experiments aimed at discovering functions of the VMPFC will be described. The first will critically examine the posited function of the VMPFC as a store for overlearned social knowledge (Milne & Grafman, 2001). The other two experiments will instead explore VMPFC

involvement in decision-making and reversal learning (Bechara, Damasio, Damasio, & Anderson, 1994; Fellows & Farah, 2003b, 2005). They will attempt to define the role of punishment in the deficits in reversal learning seen with VMPFC damage. The thesis will conclude with a synthesis of results and support for or against the speculated functions of the VMPFC.

Figure 1. VMPFC Regions

The ventromedial prefrontal cortex (VMPFC) and related areas. VMPFC consists of the medial portion of the orbitofrontal cortex (OFC) and the ventral section of the medial wall of the prefrontal cortex (PFC). Nearby regions include the frontopolar cortex and the lateral section of the OFC. Numbers indicate medial PFC and OFC subregions (Ongur, Ferry, & Price, 2003).



A. THE VMPFC: ANATOMY, CONNECTIVITY, & FUNCTION

The VMPFC is a combination of the medial subset of the region referred to as orbitofrontal cortex (OFC) (see Figure 1, lateral and medial OFC) and the ventral medial wall of the prefrontal cortex (PFC), which includes the pregenual and ventral anterior cingulate (see Figure 1, medial PFC). Nomenclature problems arise from different partitioning of the inferior prefrontal cortex in monkey versus human literature. The name VMPFC is mostly restricted to human lesion and neuroimaging studies. Monkey lesion and recording studies usually test either OFC or the medial wall. For this thesis, the VMPFC is defined as the ventral section of the

medial wall up to and including pregenual cingulate, as well as the medial half of the orbitofrontal cortex (see Figure 1, VMPFC). There are also suggestions of functional and anatomical subregions within the VMPFC.

The orbitofrontal cortex and medial PFC of the monkey have been separated into several anatomical subregions based on cortical structure (Brodmann, 1909; Carmichael & Price, 1995, 1996; Ongur et al., 2003; Walker, 1940). Recently, an extensive study on the connectivity and anatomy of these subregions (Ongur & Price, 2000) distinguished two major networks. An orbital network including orbital regions 13, 11, and 47/12 (which corresponds to monkey region 12) receives different types of food sensory input and sends projections to the caudate and putamen. A second, medial, network consisting of mostly medial regions 25, 32pl (below corpus callosum), 24b, and 10 projects to visceral control centers, caudate, putamen and nucleus accumbens. Some regions on the most ventral medial area of the PFC branch the two networks (13a, 11m, 14), providing possible integration of sensory input and visceral processing. These regions also receive input from several limbic structures, including the amygdale, perirhinal cortex and entorhinal cortex. These areas form the heart of VMPFC.

The medial and lateral distinctions of the OFC have also been separated in functional neuroimaging studies. Elliott and colleagues (Elliott, Dolan, & Frith, 2000) have succeeded in splitting some of the studies activating orbitofrontal cortex into a lateral OFC and a medial OFC region, which follows previous distinctions made by Carmichael and Price (Carmichael & Price, 1995, 1996). Elliott et al. have suggested that the OFC in general is involved in monitoring reward values (including those based on familiarity preferences). They posit that the lateral OFC is specifically involved in suppression of a previously-rewarded response, although this is based on a small amount of experiments mostly carried out by the same group. They also suggest that

medial OFC is activated when a participant is associating stimulus, response and outcome. Other imaging research also suggests a medial/lateral distinction in function (A. K. Anderson, Christoff, Stappen, Panitz, & Ghahremani, 2003), but their results suggest that the distinction is based on valence. Medial OFC (especially posterior) was found to be more active when processing pleasant odors, and was correlated with pleasantness judgements. Left lateral OFC was found to be involved in response to unpleasant odors, and was correlated with unpleasantness ratings. However, anterior medial OFC also showed activity correlated with unpleasantness, so the medial/lateral distinction here is clouded.

Within OFC, there may also be lateralization of function. Davidson and colleagues (Davidson, 1998; Davidson & Sutton, 1995; Henriques & Davidson, 1991) associate right frontal cortex with withdrawal-related behavior and left frontal cortex with approach-related behavior. Others specify the right hemisphere as critical for the representation of emotion (Borod et al., 1998). The distinction of approach-based emotions in the left hemisphere was partially supported in a recent whole brain meta-analysis by Murphy and colleagues (Murphy, Nimmo-Smith, & Lawrence, 2003). Specifically, happiness and anger activations were expressed more in the left frontal lobe than the right.

Another possible subspecialty is that of posterior VMPFC, which several have shown in monkeys to be innervated primarily from limbic and paralimbic emotional-related structures such as the amygdala and ventral striatum (Carmichael & Price, 1995; Ledoux & Muller, 1997). This emotional connection is supported by monkey lesion literature, showing taming or emotional blunting resulting from damage to the posterior VMPFC (Damasio & Van Hoesen, 1983). Note, however, that recent human anatomical comparisons suggest that subgenual posterior VMPFC in humans corresponds to more anterior regions in monkeys (Ongur et al.,

2003), so the correspondence to posterior human VMPFC may be questionable. Human subgenual cingulate has been closely related to mood disorders, especially depression (Goldapple et al., 2004). Ongur and colleagues (Ongur, Drevets, & Price, 1998) have found glial cell reductions in subgenual cingulate in mood disorders. In addition, right posterior VMPFC was active during provocation of anxiety in OCD, PTSD and phobic patients (Rauch, Savage, Alpert, Fischman, & Jenike, 1997).

Portions of the VMPFC are also implicated in emotion processing and mood disorders. The VMPFC includes limbic regions on the medial wall, including subgenual and pregenual cingulate. The subgenual cingulate, as noted above, has been especially implicated in mood disorders (Ongur et al., 1998), mood processing (Mayberg et al., 1999) and mood changes (Goldapple et al., 2004). A meta-analysis of imaging studies of emotion by Wheeler & Siegle (submitted, 2006) has also indicated that the *pregenual* cingulate may have a connection to processing emotional stimuli, especially secondary inducers of emotion such as reflection or imaging of emotional events.

In summary, the VMPFC is an area including both orbital and medial frontal subregions. Functional differences suggest that the VMPFC (medial OFC portion) is distinct from lateral OFC. Also, there may exist further functional subregions within the VMPFC, left vs. right and posterior vs. anterior. The VMPFC has connections which implicates it in the integration of sensory input and visceral processing, as well as emotion processing.

B. BRIEF INTRODUCTION TO HYPOTHESES OF VMPFC FUNCTION

Several hypotheses posit different roles for VMPFC function. These will be tested and examined in the three experiments of this thesis. As the evidence and explanation for these

theories will be expanded in the experiment chapters, only a brief review of the theories is presented here as a reference.

1. SOCIAL KNOWLEDGE HYPOTHESIS

Milne and Grafman (Milne & Grafman, 2001) posit that the VMPFC is involved in automatically accessing overlearned social knowledge. For example, overlearned social knowledge would consist of social stereotypes (gender, race attributes) that is culturally ingrained. We have learned in recent years that an automatic, not conscious activation of social knowledge of stereotypes can bias responses (Blair & Banaji, 1996; Greenwald, McGhee, & Schwartz, 1998). The automatic access of the stereotypes may occur, for instance, when interviewing a job applicant. Even those who believe, or show through self-report, that they do not hold stereotypes may show automatic implicit stereotyping effects (Devine, 1989). Milne and Grafman (Milne & Grafman, 2001) have tested this implicit social knowledge in brain-damaged patients. They have found abnormal performance on the implicit social test in VMPFC patients, as compared to patient controls and normal controls. Based on their results, they have suggested that the VMPFC is involved in automatically accessing over learned social knowledge.

2. SOMATIC MARKER HYPOTHESIS

An alternative, and widely acknowledged hypothesis of VMPFC function, involves the Somatic Marker Hypothesis, put forward by Damasio (Damasio, 1994). In this hypothesis, emotional changes (somatic states) are expressed mainly by changes in the representation of body state (skin conductance, heart rate, etc.). As regards VMPFC function, the hypothesis asserts that the VMPFC links situational contexts to the body state changes (emotion) relevant to those contexts, based on previous emotional learning (Bechara, Tranel, & Damasio, 2000). For example, in a dark street in which you have previously been physically assaulted, the VMPFC

would use past arousal-situation links to access the situation-relevant somatic state equivalent to the emotion of anxiety (sweating, heart racing, etc.). This hypothesis reinterprets the social deficits in VMPFC patients as an affective disorder. In a situation that should induce embarrassment, for example, VMPFC patients are unable to engage this previously learned emotion, and therefore behave in socially inappropriate ways.

3. FLEXIBLE STIMULUS-REINFORCER LEARNING HYPOTHESIS

In a third group of hypotheses, several researchers view the OFC (including VMPFC) as critical to reward processing (Gaffan, Murray, & Fabre-Thorpe, 1993; Rolls, 1996, 2000; Schultz, Tremblay, & Hollerman, 1998, 2000). Rolls suggests that the OFC is involved in rapid stimulus-reinforcer learning and reversal (Rolls, 1996), and acts in the maintenance of reinforcement associations of large numbers of stimuli for long periods of time. The proposed role in reward processing and reversal is supported by work done by both monkeys and humans which suggests that VMPFC or OFC damage results in reversal learning deficits (Fellows & Farah, 2003b, 2005; Iversen & Mishkin, 1970; Rolls, 1996).

Although research exists which may be viewed as support for all three of the above hypotheses, it is also possible that the social and emotional deficits due to VMPFC damage may be reinterpreted as due to simple underlying deficits in stimulus-reinforcer processing.

II. EXPERIMENT 1: TESTING SOCIAL KNOWLEDGE THEORY

The VMPFC has been shown to be involved in processing stimuli with reinforcing and/or affective properties and appears to play an important role in certain forms of flexible reinforcement learning (Bechara et al., 1994; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001; Rolls, 2000). However, separate lines of research have also implicated VMPFC in social processes (Milne & Grafman, 2001). Current research leaves open the question of whether the social involvement of VMPFC may be due to its involvement in more basic processing such as stimulus-reinforcer associations.

A few lines of research suggest that the VMPFC is involved in socially related processing. For example, some patients with VMPFC damage have severe disorders in interacting appropriately in social contexts, but the processes underlying these changes remain unclear (Barrash, Tranel, & Anderson, 2000). Similarly, on a social behavior questionnaire completed by informants, VMPFC patients were rated as less socially adept (including empathy ratings, noting others' moods, and impulsivity) (Hornak et al., 2003). Interestingly, processing morally unpleasant stimuli such as physical assaults has been shown to activate VMPFC in normal subjects studied with fMRI (Moll et al., 2002), as has viewing tapes of actors whose motor expressions of emotion do not match the emotion of their speeches (expressions were happy when story was sad) (Decety & Chaminade, 2003). These types of stimuli can be seen as socially relevant, in that they typify socially unpleasant situations (viewing assault or being lied to). In monkeys, lesions of the OFC can lead to a social disorder, specifically, more aggression when a human intrudes on their territory (Izquierdo & Murray, 2005). These studies point to the

possible involvement of VMPFC in different social functions. Milne and Grafman (Milne & Grafman, 2001) have proposed a specific hypothesis in this regard: that the VMPFC holds representations of social knowledge.

The Milne & Grafman theory arises from the social cognitive neuroscience arena, and the testing of social stereotypes. It has long been known that stereotypes can be revealed explicitly, through questionnaires. However, we have learned in recent years that social knowledge of stereotypes can be revealed implicitly, through such tasks as the IAT (Implicit Association Task: (Greenwald et al., 1998) or by stereotype priming (Blair & Banaji, 1996). For example, the IAT tests the implicit links between stereotype categories (such as "American") and stereotype-associated categories (such as "good"). Participants are asked to categorize exemplars (e.g., "Babe Ruth", "flower") from each category, and in critical trials certain categories result in the same button press (press 1 if example is either "good" or "American", and 2 if it is "bad" or "foreign"). The combined categories can be stereotypic combinations or counter-stereotypic combinations. In normal adults, when the combined categories are similar or associated stereotypically, responses to the exemplars are faster than when the concepts don't match. Even those who believe, or show through self-report, that they do not hold these stereotypes may show automatic implicit decision-time stereotyping effects (Devine, 1989).

In an IAT of social (gender) stereotypes, Milne and Grafman (Milne & Grafman, 2001) found that patients with damage to the VMPFC showed a much lower IAT effect compared to normal controls and patients with frontal damage outside the VMPFC. Specifically, in VMPFC patients, they found that reaction times to the stereotyped group categories (male & strong or female & weak) were not faster than reaction times to the counter-stereotypical category groupings (male & weak or female & strong). They suggest that this implicates the VMPFC in

the representation of social knowledge. However, this study does not strongly support the VMPFC as a selective processor of social implicit knowledge. First, the patients were separated into VMPFC versus non-VMPFC groups post-hoc, based on their performance on the test. This can lead to spurious lesion-location to behavior associations, particularly when sample size is small; there were only three patients in the non-VMPFC group (7 in the VMPFC group) in the Milne & Grafman study. Selecting groups a priori to test a pre-specified brain structure-function hypothesis is an inferentially stronger approach. Second, the Milne & Grafman study had no control task which tested non-social knowledge, providing no grounds to judge whether the observed difference in the VMPFC group was specific to social knowledge. Third, the methods of design and analysis of the IAT that were available at the time have since been shown to run the risk of producing spuriously large IAT scores when responders are slow (as they were for non-VMPFC patients in Milne & Grafman, 2001) (Greenwald, Nosek, & Banaji, 2003).

Experiment 1 critically examines all 3 of these issues. Patients with frontal lobe damage were assigned a priori to either VMPFC or non-VMPFC damaged groups, and tested on both the social IAT and an additional *non-social* IAT. Task design and analysis followed the newer methods that minimize potential confounding effects from non-specific slowing. If VMPFC does play a crucial role in processing social knowledge, the original finding of a lower IAT in the social task should be replicated in this new group of patients, and their performance on the non-social IAT should be normal.

A. GENERAL PARTICIPANTS

Procedures were approved by the Institutional Review Board at the University of Pennsylvania. All participants provided written, informed consent prior to participation in the study, in accordance with the declaration of Helsinki and were paid a nominal fee for their time.

Recruitment was as follows: Patients with focal brain injury were drawn from the University of Pennsylvania Center for Cognitive Neuroscience lesion database. Potential patients were identified by their attending physician, who obtained permission from the patient for the investigator to discuss the proposed research with them. All patients with damage primarily involving the frontal lobes anterior to the precentral sulcus, and without other neurological, medical or psychiatric conditions likely to affect cognition were eligible. Normal controls matched for age and education were drawn from the University of Pennsylvania Center for Cognitive Neuroscience normal database and from the community at large, recruited through newspaper advertisements and posters. For the second set of experiments, one patient and several controls were also recruited and run through colleagues at the Montreal Neurological Institute.

Patient lesions were traced from MR or CT images onto the standard Montreal Neurological Institute brain using MRIcro software (Rorden & Brett, 2000) by a neurologist experienced in imaging interpretation. Out of all eleven nonVMPFC patients, seven lesions involved the lateral portion of one hemisphere (four to the left and three to the right hemisphere). The remaining four nonVMPFC lesions were dorsomedial (two lateralized to the left and two to the right hemisphere). VMPFC damage was either definitely or probably bilateral in all cases, although asymmetrically so in many (see Figure 2). Patients were tested at least 6 months after brain injury had occurred.

Controls were chosen to match the patient demographics (age and education), and were required to be right-handed, have English as their mother tongue, normal vision, and full use of their hands. Controls were not taking psychoactive medication and had no history of psychoactive medication use nor psychiatric, neurologic, or medical disorders likely to affect

cognition. Controls were excluded if they scored <28/30 on the mini-mental status examination (MMSE) (Folstein, Robins, & Helzer, 1983).

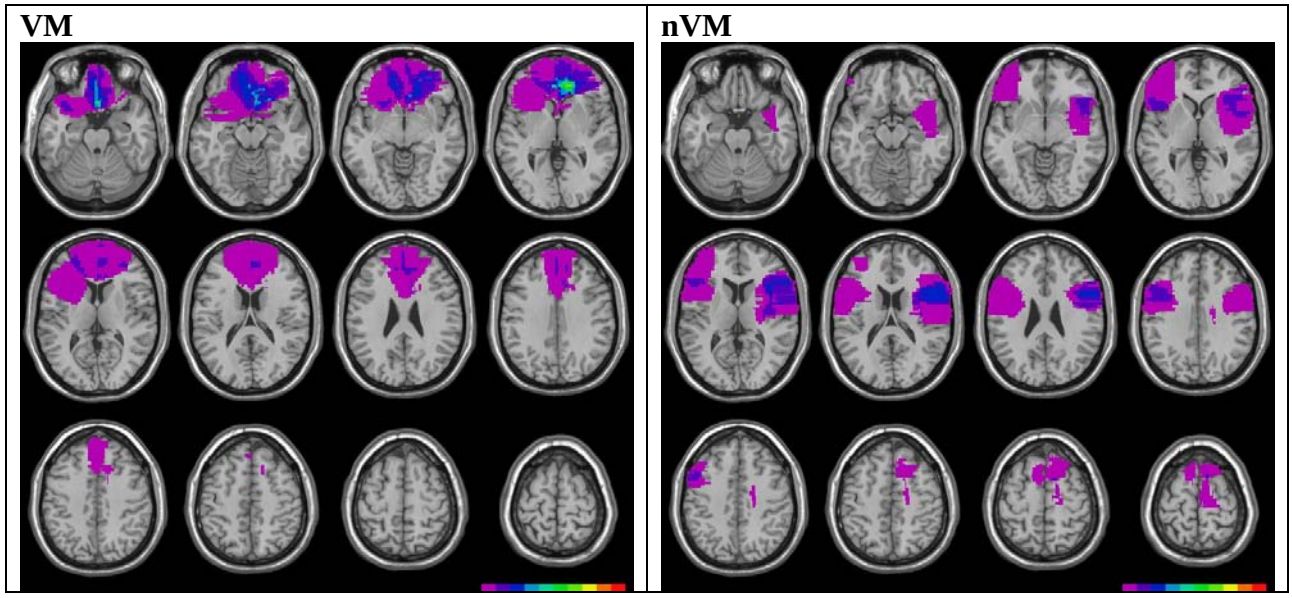
Participants who completed all studies were run in two sessions, with the social knowledge tasks completed in the first session. This first session lasted about 2 hours, and included the social and nonsocial implicit tasks, the stereotype questionnaire, and an unrelated questionnaire. The general ordering of the tasks was as follows: an implicit task, the unrelated (mood) questionnaire, the second implicit task, and the gender stereotype questionnaire. The order of the two types of implicit tasks was counterbalanced within groups. In the second session, the gambling task, probability learning task, a risk-taking task and several questionnaires were run (to be described in Chapters III and IV). The general ordering of the second set of tasks was as follows: the gambling task, the unrelated (mood) questionnaire, the probabilistic learning task, the risk-taking task, a behavioral inhibition/activation questionnaire, and the gambling questionnaire. The ordering of the gambling task and probabilistic learning task was counterbalanced within groups.

B. PARTICIPANTS

Ten patients with ventromedial prefrontal (VM) damage, 10 patients with frontal damage outside of the VMPFC (nVM), and 16 normal controls (C), matched for age and education with the frontal groups, participated in the IAT study. The area and overlap of the lesions in the two frontal groups is shown in Figure 2. There was no significant difference in lesion volume between the patient groups (Mann-Whitney U test, $p=0.50$).

Figure 2. Lesion extent and overlap for ventromedial (VM) and non-ventromedial (nVM) prefrontal groups for IAT experiments.

Color bar at bottom represents the degree to which damage was common across patients. Purple indicates areas damaged in one patient, and every increase in color corresponds to the addition of one other patient. Coordinates are radiological (L side = R hemisphere).



C. DESIGN

The method used was modeled on the most recent design for the IAT as described in (Greenwald et al., 2003). Each IAT consisted of 242 total trials. Condition 1 required subjects to discriminate between male names and female names (social) or between insects and flowers (nonsocial). Conditions 2 and 4 required subjects to discriminate words from one of two stereotypical attributes. Conditions 3 and 5 combined stimuli that were used in conditions 1, 2, and 4. These combination conditions involved mapping either a stereotypically associated attribute (e.g., female + weak, male + strong) or a stereotypically unrelated attribute (female + strong, male + weak) to the same hand. The IAT effect comes from a calculation of the difference in response times between conditions 3 and 5 (congruent and incongruent category groupings). If there is an implicit association between the target categories, then the subject

should find it easier to map the two categories together, which will be reflected in shorter response times. For example, if the participant associates women with weak and men with strong, then the associated response times will be faster, for example, when the participant is mapping female names and weak words to the same response key than when the participant is required to map females names with strong words. The organization of conditions used in the experiment is shown in Table 1.

Table 1. Implicit Association Task Designs

The (Greenwald et al., 2003) design used is shown. The experimental block order, number of trials, condition type, response key, and stimulus types for the social and nonsocial versions of the IAT are shown. Across participants, stereotypical and nonstereotypical groupings were counterbalanced (attribute types switched left to right). Response key was also counterbalanced across participants.

			Social IAT		Non social IAT	
Block	**# trials**	**Condition**	**Left key**	**Right key**	**Left key**	**Right key**
1	30	1	Females	Males	Insects	Flowers
2	30	2	Weak	Strong	Unpleasant	Pleasant
3	16	3 Practice	Weak/Females	Strong/Males	Unpleasant/Insects	Pleasant/Flowers
4	60	3 Test	Weak/Females	Strong/Males	Unpleasant/Insects	Pleasant/Flowers
5	30	4	Strong	Weak	Pleasant	Unpleasant
6	16	5 Practice	Strong/Females	Weak/Males	Pleasant/Insects	Unpleasant/Flowers
7	60	5 Test	Strong/Females	Weak/Males	Pleasant/Insects	Unpleasant/Flowers

The older design used by (Milne & Grafman, 2001) uses five blocks, three for single category matching, and two for combination categorizations. The new design (Greenwald et al., 2003) employed here utilizes combination practice trials before the combined category blocks (Practice for Conditions 3 and 5) to help minimize order effects (See Table 1). The order of congruent or incongruent trials was counterbalanced across participants for each group (controls, non-VMPFC, VMPFC) using the same method as (Milne & Grafman, 2001). Response key was additionally counterbalanced across participants for each group (left and right balanced). The number of trials for all blocks except the practice blocks was also increased (single categorizations from 20 to 40 and double categorizations from 40 to 60) in order to maximize

power. The combined category practice blocks had 16 trials instead of the 20 used in the Greenwald et al. design (Greenwald et al., 2003).

D. STIMULI

The stimuli for the social IAT consisted of the 15 male names, 15 female names, 15 “weak” words, and 15 “strong” words used by (Milne & Grafman, 2001) based on the work of Rudman and Kilianski (Rudman & Kilianski, 2000). The stimuli for the nonsocial IAT (see Appendix A) were a subset of those used by (Greenwald et al., 1998) including insect names, flower names, pleasant words and unpleasant words. All stimuli are provided in Appendix A.

E. PROCEDURE

At the beginning of the experiment, participants were shown and read the following instructions:

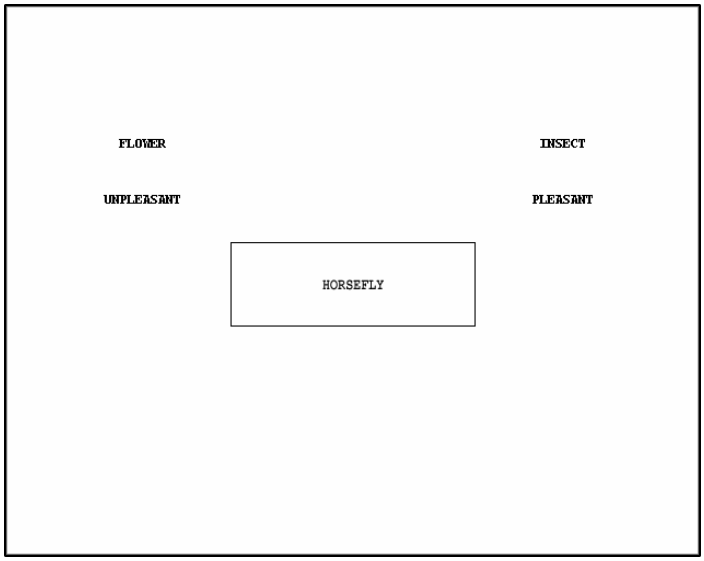
> "In this experiment, you will be categorizing words. Words will appear in a box in the center of the screen. Two or four categories will appear to the left and right of the box. You will press the left key (index finger on "z") if the word in the box best fits either category on the left, and vice versa for the right (index finger on "/")."

Participants were then walked through a sample trial, two slow example trials, five practice trials with two categories, and four practice trials with four categories. If the participant was still confused or not responding correctly, they were run through the instructions and practice trials again. Before starting the experimental trials, they were told to respond as quickly and accurately as possible. Comprehension of the task was tested throughout training by probing understanding and by prompting for questions.

At the start of each block, the experimenter would first show the categories and tell the participant which key to press for which category. Participants viewed stimuli in black bold 18 point uppercase type on a white background of a computer monitor. During each block, a rectangle was present on center screen at all times, and on either side of it were the category titles to remind subjects which side to press to make the correct response. In the rectangle, instances of each category would appear. Each instance would remain on the screen until one of the two keyboard buttons was pressed. A representative view is displayed in Figure 3. If the response was incorrect, a red "X" was shown in the rectangle. After each block, the participant was given a brief (about 1 min.) break. Between social and nonsocial IAT administration, participants were given longer breaks and were given an unrelated questionnaire (total duration about 5-10 min.). On finishing the second IAT experiment, participants were administered "The Ambivalent Sexism Inventory" (ASI) (Glick & Fiske, 1996), which measured explicit gender attitudes.

Figure 3. Example of view for a nonsocial incongruent combination trial of the IAT.

The four categories appear to the upper left and right of the center. the target stimulus appears in the box. The participant presses the left key if the target is a flower or an unpleasant word, and the right key if the target is an insect or a pleasant word. Normal controls will be slower to respond in this case, as the combined categories are incongruent.



F. RESULTS

Categorization performance was high for all three groups. Normal controls, VMPFC-damaged patients and non-VMPFC frontal damaged patients all performed above 90% accuracy on single categorizations (flower/insect, pleasant/unpleasant, male/female, and weak/strong).

The VMPFC deficit in implicit associations and specificity of that deficit to social knowledge was examined with a two-way mixed factor ANOVA on IAT Effect (using D^4 measure from Greenwald et al., 2003) with a between-subjects factor of Group (C, nVM, VM) and a within-subject factor of Knowledge Type (Social, Nonsocial). One outlier (case with value between 1.5 and 3 interquartile ranges from either the 75th quartile or the 25th quartile of the group and more than 1.5 standard deviations from the mean group score) was identified in the VM group and excluded from analysis. There was a significant main effect of Group [$F(2,32)=7.22$, $p < 0.005$)]. As shown in Figure 4, VM patients showed lower implicit association effects across social and nonsocial knowledge (M = .15, SE = .10) compared to both the normal control group (M = .62, SE = .08) and the nVM patients (M = .535, SE = .10), Bonferroni tests, $p<.05$. Critically, there was no significant interaction between Group and Knowledge Type ($p=.91$), revealing that the Group effect did not differ between Social and Nonsocial Knowledge IAT Effects (see Figure 4). Raw data (adjusted reaction time data from step 9 of Greenwald et al., 2003) is provided in Table 2.

Further one-way ANOVAs separated by knowledge types clarified the quality of the two-way lack of interaction. For the Social Knowledge as well as for the Nonsocial Knowledge ANOVA, significant group effects were found (S, $F(2,34) = 4.76$; NS, $F(2,34) = 4.89$; both $p<.05$). Post-hoc tests revealed that VM patients had significantly lower IAT Effects than normal controls (Bonferroni, $p<.05$) for both types of knowledge. VM patients also had numerically lower IAT Effects than nVM patients for both types of knowledge, but this did not

reach significance in the one-way ANOVA tests. Normal controls performed indistinguishably from nVM patients in both one-way ANOVA tests.

Figure 4. IAT Effects for Knowledge Type by Group.

There is no significant interaction for IAT Effect between Group and Knowledge Type. IAT Effect was lower for the VMPFC group (VM) than for both control groups (nVM=Non-VMPFC frontal patients, C=Normal controls). Critically, this was true for both social (S) and nonsocial (NS) knowledge types. Bars show 95% confidence intervals.

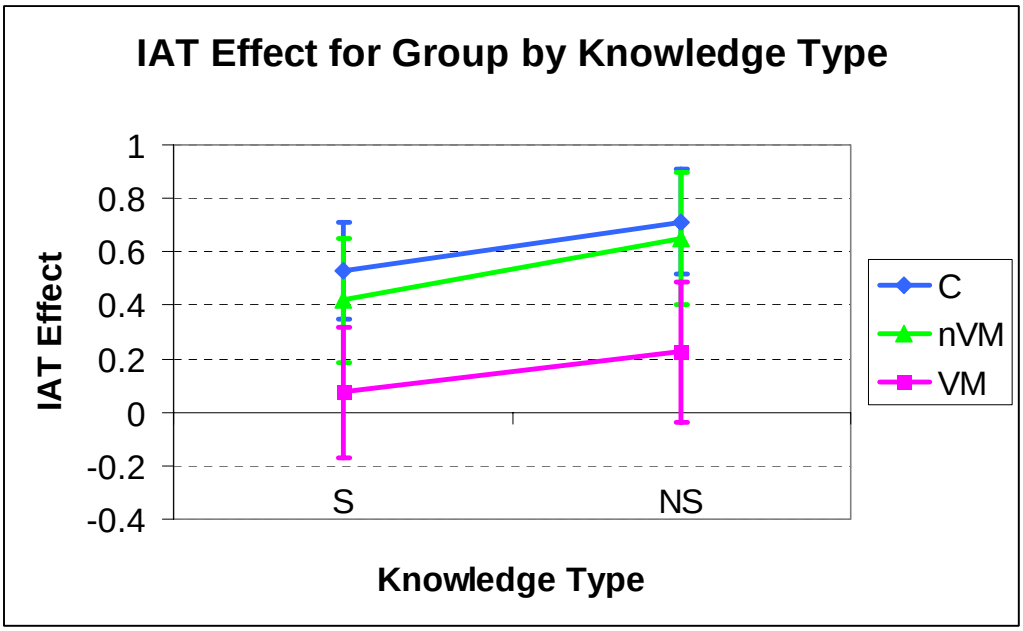


Table 2. IAT Adjusted reaction time data.

VM patients showed overall slower reaction times, and on average participants were slower for the nonsocial knowledge task. The improvement in reaction time on congruent trials does not show any noticeable interaction between group and knowledge type. Raw reaction time data is adjusted according to Steps 1- 9 in Greenwald et al. (2003).

Raw Data (Adjusted[a] Mean Reaction Times)

	Nonsocial		Social	
	Congruent	Incongruent	Congruent	Incongruent
C	961.82	1190.00	906.70	1063.08
(SD)	*(263.14)*	*(228.99)*	*(189.32)*	*(252.46)*
nVM	979.59	1223.85	951.86	1083.71
(SD)	*(233.46)*	*(197.13)*	*(196.29)*	*(186.65)*
VM	1620.19	1893.63	1501.08	1637.87
(SD)	*(576.97)*	*(500.96)*	*(450.23)*	*(432.10)*

a. Greenwald et al., 2003; Step 9

There was no correlation between lesion volume and either Social (Rho=-0.14, p=0.55) or Nonsocial (Rho=.17, p=0.47) IAT Effect. This provides further support for the claim that these findings are specifically related to the location of the brain injury, but not to the extent of the injury.

G. DISCUSSION

The results support two major conclusions. First, they replicate the findings of (Milne & Grafman, 2001) in a new group of patients, and using optimal methods for measuring IAT. VMPFC damage was associated with a lower implicit association effect compared to both healthy controls and to patients with frontal damage outside VMPFC. However, this lower IAT effect in VMPFC patients was not specific to social knowledge: These patients also demonstrated a smaller IAT effect with non-social stimuli. These findings argue against the claim that the VMPFC is selectively involved in processing implicit social knowledge.

The reduction in IAT effect in VMPFC patients, therefore, appears to be a more general effect, one not specific to social knowledge. What, then, may produce this change? Possible explanations include: a deficit in semantic knowledge, a deficit in managing response conflict, or a general deficit in linking reinforcement or affective value to stimuli.

Could the minimization of the IAT effect in VMPFC patients be due to involvement of damaged regions in semantic processing? Semantic processing has been widely associated with the left inferior prefrontal cortex (LIPC) (Buckner, Raichle, Miezin, & Petersen, 1996; Fiez, 1997; Gabrieli, Poldrack, & Desmond, 1998; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001), typically BA 45 and 46 (although in some studies extending more ventrally to BA 47) Some

VMF subjects had damage extending into BA 45 and BA 46, but it was more commonly and more extensively damaged in the non VMPFC group, in whom no change in IAT was found. (mean voxel damage to BA 45 & 46: nVM=2824.9, VM=1315.4). Performance was also high for categorization in all three groups (above 90% for single categorizations). This makes it unlikely that the difference in VMPFC patients is due to damage to regions classically involved in semantic processing.

Similarly, a deficit in response conflict performance is also an unlikely explanation for the low VMPFC IAT effect. The IAT effect is largely due to response conflict during incongruent trials, in which target words (e.g., "wasp") produce conflict in response (press left for insect/pleasant, or right for flower/unpleasant) due to the implicit associations. Although response conflict has been increasingly associated with anterior cingulate activity (Carter et al., 2000; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001), the usual prediction is that damage to these areas would lead to exaggerated response conflict. Specifically, a problem with response conflict would lead to poorer performance on the incongruent trials than the congruent trials, leading to a *larger* IAT Effect for VMPFC patients. This is the opposite effect of that seen in the current study. Further, only one VMPFC patients had damage extending into dorsal ACC.

An alternative explanation for the IAT difference lies in the role of VMPFC in linking stimuli to affective value. Various lines of research have emphasized the role of VMPFC in associative learning and decision making when reinforcing or emotionally-valenced information is involved (Bechara et al., 1994; Rolls, Critchley, Mason, & Wakeman, 1996). The attributes in the IAT experiments, whether social or non-social, are valenced (pleasant/unpleasant; weak/strong). It may be that the reduced IAT effect in both social and non-social versions of the task in VMPFC patients is due to a general impairment in representing the (learned) valence of

stimuli. However, no difference between processing negative (unpleasant) versus positive (pleasant) exemplars was found in either accuracy or reaction times for the VMPFC patients compared to the other groups. This suggests that if there is a difference in reinforcement or valence decisions, it may be a more complex function. *Flexible* associations between valence or reinforcement and stimulus choice, and not simple processing of valenced stimuli, may better explain deficits resulting from VMPFC damage (Rolls, 2004). Specifically, the VMPFC change in IAT Effect may reflect differences in flexible associations between valenced attribute categories (e.g., pleasant) and group categories (e.g., flower). The valenced categories are associated with different group categories depending on whether the block is congruent or incongruent. VMPFC-damaged patients may show lower IAT Effects because of deficits in processing this switch in valence association.

In conclusion, the results show agreement with the (Milne & Grafman, 2001) data, in that VMPFC patients have a lower implicit association effect than normal and patient controls. This agreement held up when groups were divided a priori, a larger control patient group was studied, and new methods were used which reduce possible spurious IAT effects. However, the IAT difference in VMPFC patients was not specific to social knowledge, suggesting that the VMPFC does not play a special role in representing social knowledge. Rather, these findings support a more general role for VMPFC in flexibly linking affective or reinforcement value to stimuli. This interpretation links the IAT findings to a broader set of results in both animal and human work showing that VMPFC is important in representing the current reinforcement value of primary reinforcers (Rolls, 2000; Rolls, Critchley, Browning, Hernadi, & Lenard, 1999) or conditioned reinforcers (Schoenbaum, Chiba, & Gallagher, 1998; Schultz et al., 2000). Such a general role for VMFPC also links these findings to the literature on the role of VMPFC in

decision making, where it has also been argued that VMPFC function involves using affective signals to guide behavior (Bechara et al., 2000).

III. EXPERIMENT 2: ROLE OF PUNISHMENT IN REVERSAL LEARNING DEFICIT

One of the leading theories on VMPFC function involves the Somatic Marker Hypothesis, put forward by (Damasio, 1994). This hypothesis was advanced to explain impaired performance in a gambling task (Iowa gambling task; IGT) by VMPFC patients. The IGT was in turn intended to capture the difficulties these patients had with everyday decisions. In the IGT (Bechara et al., 1994), the participant chooses between decks that give higher rewards but which also have severe punishments, versus decks that give lower rewards but also have much smaller punishments. The latter, the "good" decks, are eventually preferred by most normal participants. However, the VMPFC patients did not show the same aversion to the "bad" decks. VMPFC patients also did not show a preceding skin conductance change when picking from the "bad" decks. (Damasio, 1994) concluded that VMPFC is involved in linking a situation with an associated somatic state (e.g., skin conductance change) through learning.

However, recent research suggests an alternative explanation of the (Bechara et al., 1994) results. Researchers have posited that the deficit in the Iowa gambling task is due to implicit reversal learning in the task (Fellows & Farah, 2005). The experiment is set up so that the "bad" decks pay out rewards for a time before the large punishments are given, biasing participants to the "bad" decks at the beginning of the experiment, and requiring a reversal to the "good" decks in order to achieve "normal" performance. Fellows and Farah (Fellows & Farah, 2005) tested patients in a non-reversal version of the Iowa gamble, and revealed that the impairment of the VMPFC patients was reduced, showing support for a more general impairment in stimulus-reinforcer reversal.

The hypothesis that the VMPFC, or, more particularly, the orbitofrontal cortex (OFC), is involved in rapid stimulus-reinforcer learning and reversal is also espoused by Rolls (Rolls, 1996). Rolls posited that the OFC is involved in maintenance of reinforcement associations of large numbers of stimuli for long periods of time. This reward-specific function of the OFC was revealed in tasks in which monkeys were fed a specific food to satiety. The majority of OFC olfactory neurons display decreased responses to the food when the monkey is satiated with that food, as opposed to food in which the animal was not satiated (Critchley & Rolls, 1996). The proposed role in reward processing and reversal is supported by work which showed that medial OFC (VMPFC) lesions in monkeys resulted in an impairment of the ability to reverse visual stimulus reinforcer relationships (Iversen & Mishkin, 1970). Specifically, some monkey OFC neurons respond to non-rewards when an expected reward is not obtained, and a reversal is required (Thorpe, Rolls, & Maddison, 1983). Not only in monkeys, but also in humans, the VMPFC has been linked to reversal learning and reinforcement. An fMRI study (O'Doherty, Critchley, Deichmann, & Dolan, 2003) revealed that VMPFC response to feedback during reversal learning differed depending on whether the feedback signaled the need for a reversal. Also, several studies of patients with damaged to VMPFC (Fellows & Farah, 2003b, 2005) have provided evidence that it is critical for the ability to flexibly process stimulus-reinforcer associations.

However, it is not clear whether the impairment of VMPFC patients is specific to situations which involve punishment (such as the Iowa gambling task), or whether the impairment is general, including situations involving any reversal of stimulus-reinforcer associations (e.g., those which involve only reward). There is some indication that punishment is processed differently from reward. First of all, punishment is weighted more strongly than

reward (Kahneman & Tversky, 1991): a little goes a long way. Second, in a classification learning task using fMRI, the VMPFC has been found to be selectively responsive to negative over positive feedback (Aron et al., 2004). If the VMPFC is more critical for situations involving negative feedback, there may be a sparing of reversal abilities in a rewards-only situation. Personal communication from Fellows reports that VMPFC patients were also normal at a rewards only gambling task (Fellows & Farah, 2003a) that did *not* include reversal (rewards-only w/o reversal). The following experiment will examine performance of VMPFC patients in the alternative task, that involving reversal of stimulus-reinforcer associations when there is no punishment (rewards-only reversal; ROR). Integration of the current results with the data from previous gambling studies allows for specific analysis of VMPFC deficits. Normal performance on this rewards-only gamble with reversal would reveal a pattern of results (see Table 3) which supports a specific impairment in flexible stimulus-reinforcer processing only when punishment is involved (IGT task; upper left corner of Table 3). On the other hand, impairment at ROR (impairment solely on reversal tasks; left vs. right side of Table 3) would support a more general role for VMPFC in flexible stimulus-reinforcer processing, regardless of the presence of punishment.

Table 3. ROR Logic table for possible results.

Performance for VMPFC-damaged patients according to current research. Results are displayed in table according to whether the task included punishment and whether the task included an implicit reversal. Impairment at rewards-only reversal gamble indicates a general reversal deficit. Normal performance indicates a deficit specific to situations involving punishment.

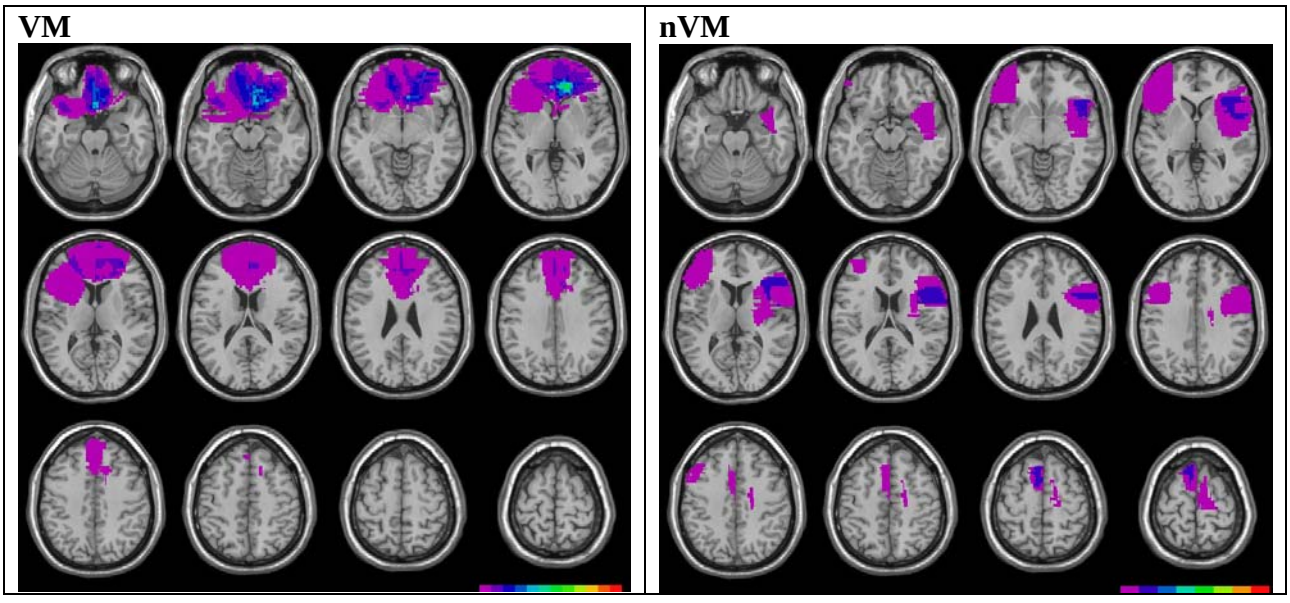
Gambling Task Performance: VMPFC-damaged patients	Reversal +	No Reversal -
Punishment +	Impaired (IGT task; Bechara et al., 1994)	OK (Fellows & Farah, 2005)
Punishment -	? (ROR; current study)	OK (Fellows & Farah, 2003a)

A. PARTICIPANTS

For the gambling task (ROR), 12 patients with ventromedial prefrontal (VM) damage, 8 patients with frontal damage outside of the VMPFC (nVM), and 21 normal controls (C) balanced for age and education were run. The area and overlap of the lesions in the two frontal groups is shown in Figure 5. There was no significant difference in lesion volume between the patient groups (Mann-Whitney U test, p=0.57).

Figure 5. Lesion extent and overlap for ventromedial (VM) and non-ventromedial (nVM) prefrontal groups for gambling task.

Color bar at bottom represents number of patients who had region overlap, where purple is one patient, and every increase in color corresponds to the addition of one other patient. Coordinates are radiological (L side = R hemisphere).

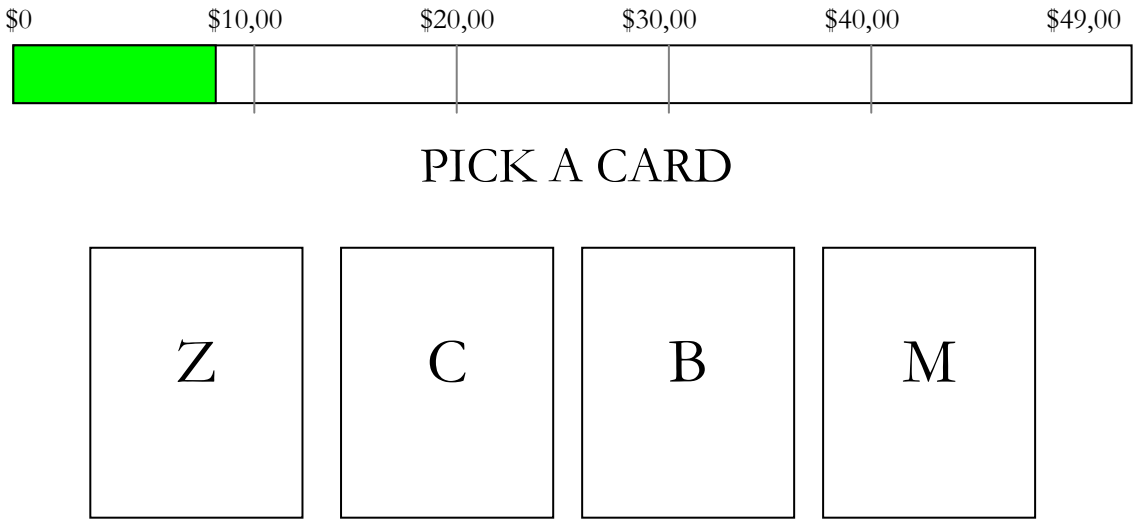


B. DESIGN, STIMULI & PROCEDURE

An adapted computerized version of the IGT was used, based on the design described in Bechara, Tranel and Damasio (Bechara et al., 2000). In the task, participants choose from four decks of cards (see adapted version in Figure 6). After each choice, the participant is given feedback about how much money was won. Some decks participants learn are "bad" decks, in that they give higher wins, but also higher losses, and are disadvantageous in the long term.

Figure 6. Rewards-only Reversal display design.

Schema of the Rewards-only Reversal task. Participants pick a card from one of four decks, and then receive a reward, which is displayed in the center of the screen after the choice. A bar at the top of the screen indicates the current amount of money won, and moves to the left as the money increases.



In the original version, some money was won on all trials, and on some trials there was also a loss. In the modified version, however, there were no losses. The original IGT includes an implicit reversal (Fellows & Farah, 2005), in that the "bad" decks start off by giving larger rewards, but then have large punishments. This version replicates this implicit reversal, by taking the net amount of money received in each trial (money won – money lost) in the original

version, and replacing that value (range: $-2330 to $170) with a positive value (range: $1 to $500). Proportional differences between the values were retained (ratio of 1: 0.2) in order to preserve the difference in goodness between the values. In this way, there were "bad" decks, but they never resulted in loss of money, only lower wins (Rewards-only reversal: ROR gamble). For example, instead of losing $2330, the participant would win money, but only $1. Instead of winning $120, they would win $490. The participants played for 100 trials, as in the original version, and the total value won was represented by a constantly present bar across the top of the screen. Task instructions were taken from Bechara et al. (Bechara et al., 2000), and modified to reflect only wins. This ROR gamble was piloted and shown to produce learning (to avoid "bad" decks) in normal adults similar to the original IGT.

C. RESULTS

The rewards only reversal performance was determined as the number of cards chosen from the "good" decks over 100 trials (Figure 8), in keeping with the most commonly used measure from the original IGT (Bechara, Damasio, Tranel, & Anderson, 1998). In addition, the number of choices from the "good" decks within each block of 20 trials is shown graphically (Figure 7), to reveal learning effects across time. A one-way analysis of variance of advantageous choices across group (C = normal controls, nVM = patient controls, VM = VMPFC damaged patients) revealed no significant difference between the groups [$F(2,38)=0.24$, $p=.79$] (see Figure 8).

Graphical description of the data (Figure 7 below) shows the poor ability to learn the "good" decks for all three groups. The similarity of performance of the VM group to the controls is consistent throughout the five blocks. Further examination of the performance reveals

that, compared to the pilot data performance from normal younger adults (66.78% chosen from good decks, and 80% for the last block), the older matched controls' performance was poor (56.57% overall, and 57% for the last block). Thus, performance not only from the patients, but also from the normal controls, was poor (group averages between 49-57%) for this task. However, although performance was poor on average, there was a significant learning effect across blocks. A two-way ANOVA of Group by Learning (Blocks 1 through 5) revealed significant learning of good decks [Main effect Learning, $F(2.75,104.5) = 4.70$, $p<.01$; Greenhouse-Geisser epsilon used to correct for non-sphericity]. Post-hoc t-tests revealed that significant learning occurred between blocks 1 and 2 ($p<.001$). No significant Group effect or interaction was seen, again supporting normal performance for VM patients.

Figure 7. ROR Choices from good decks over time.

Shows learning across all five blocks of 20 responses each for the three groups (C = normal controls, VM = ventromedial prefrontal patients, nVM = patient controls). There were no significant differences between the groups. Error bars reflect confidence intervals.

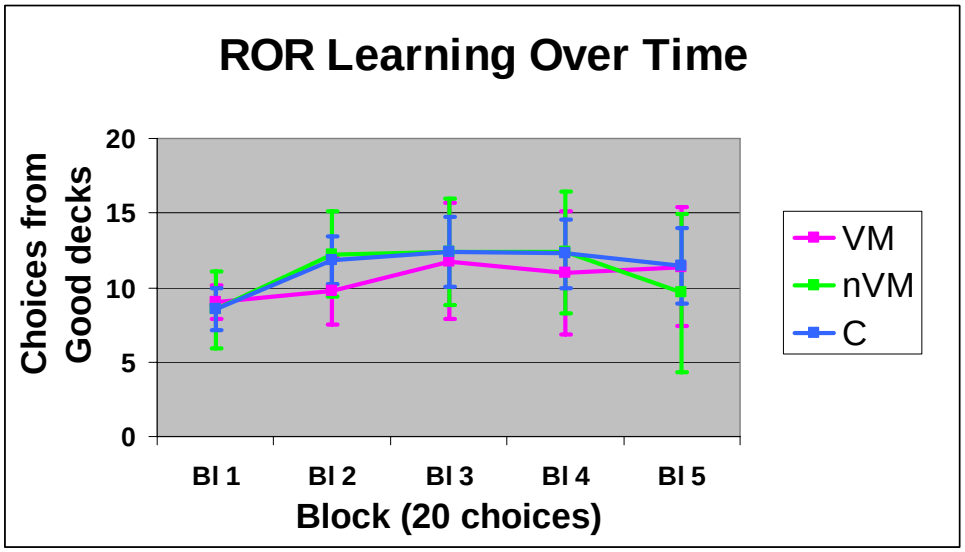
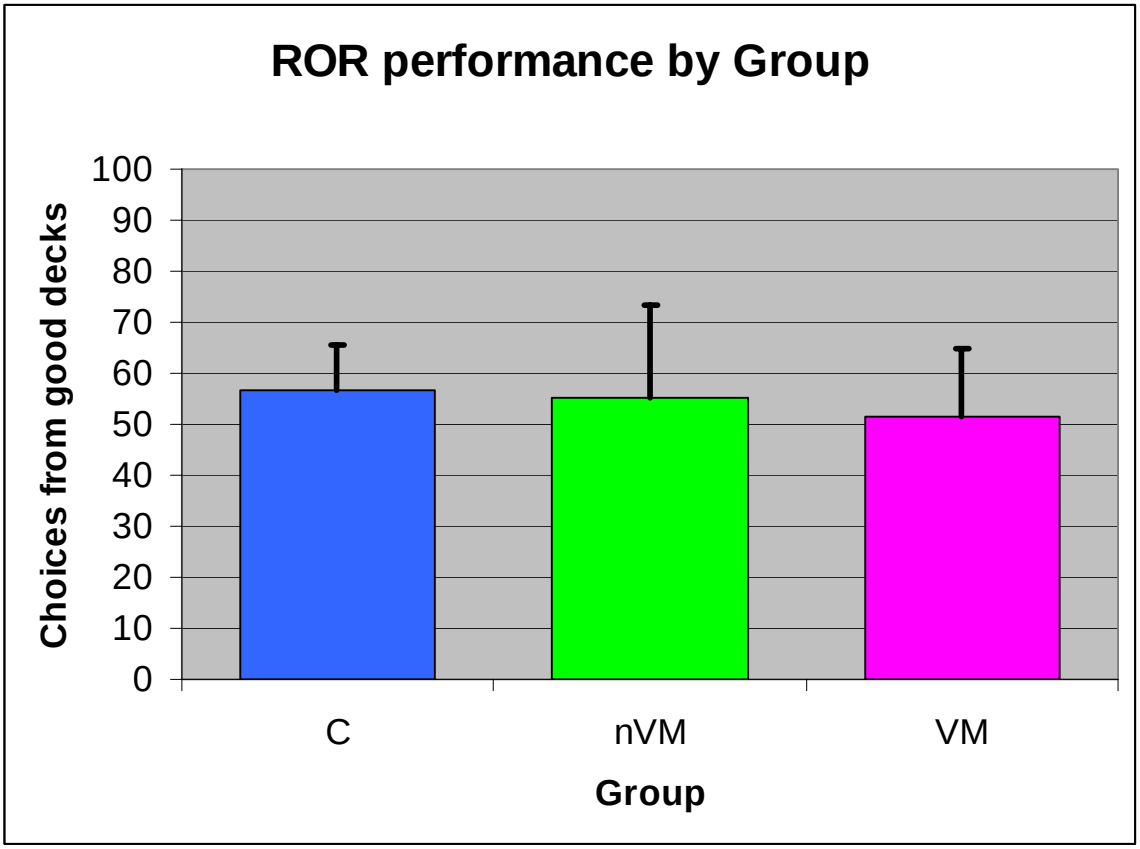


Figure 8. ROR Overall choices from good decks.

Number of choices from good decks out of 100 total trials by group. There were no significant group differences. Error bars reflect upper confidence intervals.



To eliminate the possibility that performance was influenced by the size of the lesions, a correlation was performed to examine a possible connection between total brain volume loss in the patients and ROR performance. There was no significant correlation between lesion volume and ROR performance (p=.69).

D. DISCUSSION

The results show that VMPFC performance was not impaired compared to either normal controls or patient controls on the rewards-only reversal gamble. This indicates that VMPFC patients are impaired on the gambling tasks only in cases in which both a reversal is required,

and punishment is included (see Table 4). However, this result is tempered by the absence of a strong learning pattern even amongst the normal control subjects. Nevertheless, these findings are at least consistent with the possibility that the VMPFC has a specific role in punishment processing. In support, a classification learning task using fMRI found that the VMPFC was selectively activated for negative over positive feedback (Aron et al., 2004). Then again, the VMPFC may just have more of a role in punishment processing. Considering that punishment stimuli are more salient than reward stimuli (Kahneman & Tversky, 1991), the VMPFC function may be utilized more by punishment stimuli. If the VMPFC is more critical for situations involving negative feedback, this may explain the normal reversal ability in the current rewards-only situation. A specific role of VMPFC in punishment processing may help to explain the impairments in reversal learning of VMPFC patients seen previously (Fellows & Farah, 2003b, 2005). Reversal in these cases involving loss of play money requires intact processing of punishment. For example, when a previously-rewarded stimulus is suddenly punished, it requires integration of that punishment signal in order to be able to switch responding to the alternate stimulus.

Table 4. ROR Logic table for actual results.

Performance for VMPFC-damaged patients according to current research. Results are displayed in table according to whether the task included punishment (+ or -) and whether the task included an implicit reversal (+ or -). Normal performance indicates a deficit specific to reversal situations involving punishment.

Gambling Task Performance: VMPFC-damaged patients	Reversal +	No Reversal -
Punishment +	Impaired (IGT task; Bechara et al., 1994)	OK (Fellows & Farah, 2005)
Punishment -	OK (ROR; current study)	OK (Fellows & Farah, 2003a)

Alternatively, the sparing of reversal ability in this rewards task may be due to the increase of difficulty of a task also involving punishment. In the original gambling task (IGT), the participant must integrate rewards and punishments in order to succeed. In the current task (ROR), they must only tally and compare rewards. It is possible that the VMPFC patients' difficulty arises from faulty working memory for affective stimuli (Davidson, Jackson, & Kalin, 2000). In this hypothesis, the OFC is thought to be involved in maintaining emotion during times in which no emotional stimuli are present. The hypothesis draws parallels from the role of dorsolateral prefrontal cortex (DLPFC) in working memory. Orbitofrontal cortex is thought to be involved in working memory, but of the emotional type. However, most of the evidence for this role comes from studies in which the OFC shows cue-related or expectancy activity for rewards (Hikosaka & Watanabe, 2000; Schultz et al., 2000; Watanabe, 1996). There is no evidence that the OFC is providing working memory function which involves punishment stimuli.

It is somewhat surprising that reward reversal impairments were not seen, considering the animal literature. Monkey literature has fairly consistently related OFC damage to impairment in reversal learning (visual discrimination) tasks (Butter, 1969b; Dias, Robbins, & Roberts, 1996; Iversen & Mishkin, 1970; Izquierdo, Suda, & Murray, 2004; Jones & Mishkin, 1972). These tasks involve getting a reward, or not getting a reward, which is similar to the current study. One difference is that in the monkey reversal tasks, real rewards (juice, marshmallow) were used. In the current study, play money was used, which may result in more subtle differences in response (especially when combined with a complex gambling task). It is possible that the ability to reverse in reward-only situations is actually impaired in the VMPFC patients, but that the impairment in this case is too subtle to be detected. On the other hand, there is indication that

these reward reversal deficits are localizable to the lateral OFC, not the medial OFC. Butter (Butter, 1969b) found that reversal was worst after lateral OFC lesions (than for medial OFC regions). Iversen & Mishkin (Iversen & Mishkin, 1970) also found that lateral OFC lesions resulted in worse reversal performance than medial OFC lesions, although both resulted in reversal impairment. Several of the other monkey reward reversals did not discriminate medial from lateral OFC (Dias et al., 1996; Jones & Mishkin, 1972), so it is unclear whether the deficit was due primarily to lateral OFC damage. It is possible that reward-only reversal is subserved by lateral rather than medial OFC regions, which is why this study did not find a strong impairment in VMPFC patients.

One interesting result of the current study is that normal older controls (mean age 60.6 years) did not show much preference for the good decks. This was in contrast to the performance of younger, college-age pilots performing on the same task, who developed a much stronger preference for the good decks. This may indicate a change in affective decision-making with normal aging. Reversal learning ability has been found to deteriorate with age in monkeys (J. R. Anderson, de Monte, & Kempf, 1996), although it is not clear whether this is linked to OFC changes, or changes in other brain regions.

In conclusion, VMPFC patients performed normally on the rewards-only reversal task, a modification of the Iowa Gambling Task that involves only rewards. The finding that VMPFC damage spares reward-driven learning and reversal learning, together with the published impairment of VMPFC patients on other gambling tasks (Bechara et al., 1994; Fellows & Farah, 2005), argue that VMPFC is specifically involved in the processing of punishment. However, given the limitations of the task used in this experiment, further examination of the processing of punishment and reward is needed to better support this claim.

IV. EXPERIMENT 3: LEARNING FROM NEGATIVE VS. POSITIVE FEEDBACK

Several lines of research indicate that orbitofrontal cortex is involved in reinforcement learning. First, lesion studies in primates implicate the OFC in a variety of stimulus-reinforcer settings. Lesions to the OFC resulted in impairment during reversal learning with rewards, conditioned reinforcement, and extinction of a positive reinforcer response (Butter, 1969a; Dias et al., 1996; Izquierdo et al., 2004; Jones & Mishkin, 1972; Pears, Parkinson, Hopewell, Everitt, & Roberts, 2003). Lesions to monkey OFC also result in basic changes in reinforcer processing, such as blunted fear responses to a rubber snake (Izquierdo & Murray, 2005) and less devaluation of a particular food's reinforcement value when a monkey is sated with that food (Izquierdo et al., 2004). Single cell recordings in primate OFC also support the role of ventral PFC in stimulus-reinforcer relationships. Neurons in the primate OFC are selective for preferred food items (Rolls et al., 1999), encode expected rewards and punishments (Roesch & Olson, 2004; Tremblay & Schultz, 1999), and are active after receiving reinforcement, after making errors, and when a reversal is required in reversal learning (Thorpe et al., 1983).

Lesion and fMRI studies on humans also implicate the OFC in stimulus-reinforcer processing. Damage to the OFC results in impairment in reinforcement reversal learning (Fellows & Farah, 2003b; Hornak et al., 2004) Functional MRI experiments have found that the OFC activates in different reward and punishment-related conditions. The OFC is active in response to feedback during situations in which reward or punishment is increasing (Elliott, Friston, & Dolan, 2000) as well as situations involving punishment and unexpected punishment (O'Doherty et al., 2003; O'Doherty et al., 2001). Medial OFC also responds to feedback during reversal learning (Remijnse, Nielen, Uylings, & Veltman, 2005).

Last, ERP research (Miltner, Braun, & Coles, 1997) reveals a negative deflection in ERP signal (feedback-related negativity) following feedback on incorrect performance, which may localize to the anterior cingulate (ACC). This could be considered error detection, possibly of a type critical for the ability to learn from feedback. However, there is some contention over whether ACC is the source of this activity, and if so whether it arises from ventral ACC (a portion of the VMPFC region) or more dorsal ACC (Nieuwenhuis, Slagter, von Geusau, Heslenfeld, & Holroyd, 2005) .

The evidence noted above implicates the OFC in linking stimuli to reinforcers. However, flexibly updating those links is a distinct issue. It is this 'updating' that appears to require VMPFC. Reversal learning is a classic experimental paradigm for studying this more flexible form of stimulus-reinforcer learning. In this paradigm, one must discontinue choosing a previously rewarded stimulus, and switch to a previously unrewarded stimulus. Animal and patient studies support the role of the ventromedial prefrontal cortex (VMPFC) in reversal learning. (Iversen & Mishkin, 1970) have shown that medial OFC (VMPFC) lesions in monkeys result in an impairment of the ability to reverse in visual discrimination learning. More specifically, (Thorpe et al., 1983) found that in a visual reversal learning task, some OFC neurons respond to unexpected non-rewards (during a reversal, when an expected reward is not obtained). In normal humans, O'Doherty and colleagues (O'Doherty et al., 2003) found that VMPFC response to feedback during a reversal learning task differed depending on whether the feedback was followed by a switch. (Fellows & Farah, 2003b, 2005) have extended this research to patient studies, providing evidence that the VMPFC is critical for the ability to flexibly update stimulus-reinforcer associations. However, the specific function the VMPFC serves in learning to reverse remains unclear.

The reversal of a stimulus-reinforcer association involves two critical abilities. In reversal learning, one must learn to approach a previously punished stimulus (new learning from rewards), as well as learn to reject a previously rewarded stimulus (learning from punishments). Frank and colleagues (Frank, Seeberger, & O'Reilly R, 2004) have developed a probabilistic learning task that permits the separate measurement of reward-and punishment-driven learning within the same task. This provides the opportunity to further dissect the processes underlying the reversal learning deficits observed following VMPFC damage. Perseveration during reversal learning could be due to an inability to be attracted to the previously "bad" stimulus (poor reward learning), or an inability to inhibit responding to a previously "good" stimulus (poor punishment learning), or both.

Similar questions have previously been asked in the context of the Iowa gambling task. Bechara and colleagues (Bechara et al., 2000) used variants of the original task in an effort to determine whether VMPFC patients were hypersensitive to reward, or hyposensitive to punishment. However, the complexity of the tasks resulted in an unclear result; neither hypothesis was definitively supported. The method used to test hyposensitivity to punishment (A'B'C'D' task) was to increase the amount or frequency of punishment to reward, using the same task format. There was similar, although apparently slightly better performance in the VMPFC patients, for the variant task (there was also less difference between VMPFC patients and normal controls in the variant task). However, it remains that merely increasing the punishment may not have an effect if the patients are unable to learn from punishment. The issue remains important in understanding the specific functions of VMPFC, and understanding the basis of the deficits seen following VMPFC damage in both simple and more complex learning and decision-making tasks.

This study tests whether patients with VMPFC damage are selectively impaired in either reward or punishment learning. A group of such patients were compared to a healthy control group, and to a group with frontal damage outside VMPFC. Both forms of learning was assessed using the probabilistic learning task of (Frank et al., 2004).

A. PARTICIPANTS

In total, 44 participants completed the probabilistic learning task (24 normal controls matched to the patients for age and education = C, 9 patients with frontal damage outside of the VMPFC = nVM, 11 VMPFC-damaged patients = VM). No participants had knowledge of Japanese or knew Japanese characters. A fairly high rate of task failure (unusable) data was expected based on the rate seen in Frank et al. (Frank et al., 2004), even though the current task was simplified. Of those completing the task, 35 (80%) passed the criterion used for practice trial performance, and those data (minus one VM outlier; outliers were cases with value between 1.5 and 3 interquartile ranges from either the 75th quartile or the 25th quartile of the group and more than 1.5 standard deviations from the mean group score) were used in the analyses on test performance. Those participants who passed criterion consisted of 22 normal controls (92%), six patient controls (67%) and seven VMPFC patients (64% including one outlier). Age and education were not significantly different between groups for the participants passing criterion (see Table 5). The area and overlap of the lesions in the two frontal groups is shown in Figure 9. There was no significant difference in lesion volume between the patient groups (unpaired t-test, $p=0.43$).

Table 5. Group Demographics.

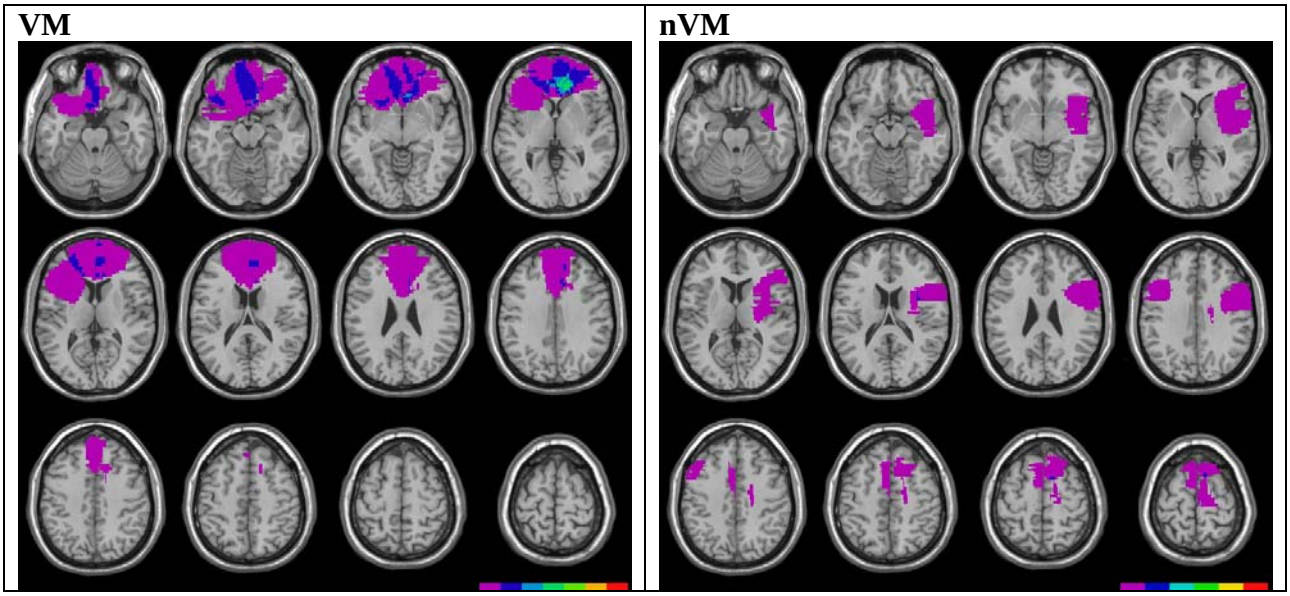
Demographics for participants passing criteria. C= Normal Controls, nVM = patient damaged frontally outside VMPFC, VM = patient damaged in VMPFC.

Demographics by Group						
	C (n=21)[a]		**nVM** (n=6)		**VM** (n=7)	
	Age	Education	Age	Education	Age	Education
Mean	61.6	13.5	50.7	15.7	53.1	12.9
SD	11.3	2.3	12.0	3.3	13.7	2.0
Min	46.0	9.0	39.0	12.0	36.0	10.0
Max	78.0	16.0	65.0	20.0	77.0	16.0

a – missing data on one control subject

Figure 9. Lesion extent and overlap for ventromedial (VM) and non-ventromedial (nVM) prefrontal groups for Probability Learning task.

Color bar at bottom represents the degree to which damage was common across patients. Purple indicates areas damaged in one patient, and every increase in color corresponds to the addition of one other patient. Coordinates are radiological (L side = R hemisphere).



B. DESIGN, STIMULI, & PROCEDURE

A modified version of the (Frank et al., 2004) probabilistic learning task was used. The task requires trial and error learning, in which the participants learn to associate some stimuli with a positive reinforcement value ("good" stimuli), and others with a negative reinforcement

value (“bad” stimuli). This is a two-alternative, forced choice task, in which participants choose one of two stimuli presented on a computer screen by pressing the left or right-side button on a keyboard. Participants sit in front of a computer monitor and view pairs of low-verbalizable visual stimuli (Japanese Hiragana characters). These figures are presented in black on a white background, in 72 pt font. Participants press keys on the left or right side of the keyboard to indicate which figure they think is “correct”. Feedback is given following each choice as a colored word displayed on the monitor (“Correct!” printed in blue or “Incorrect” printed in red). If no response is made after four seconds, the words “no response detected” are printed in red.

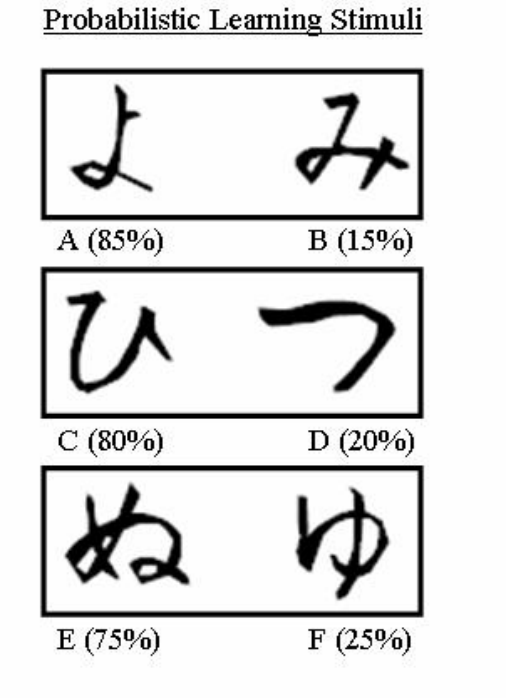
In the original probabilistic learning task (Frank et al., 2004), patients learn to favor one of a pair of stimuli in three separate sets (A vs. B, C vs. D, E vs. F) using probabilistic feedback. The first set is more consistently associated with its reward or punishment (A = reward 80%, = punishment 20%; B = reward 20%, = punishment 80%). The other sets are less stable in feedback (C vs. D: 70:30, E vs. F: 60:40). Then, in a session with no feedback, they are tested on preference for A over the stimuli not previously shown with A (C, D, E, F), which tests how well they learned to favor A, or how well they learned from positive feedback on A. They are also tested on preference for B versus novel pairings, which tests how well they learned to reject B, or how well they learned from negative feedback. In the training sessions, three different stimulus pairs (AB, CD, EF) are presented in random order. Probabilistic feedback follows the choice to indicate whether it was correct or incorrect, with the probabilities determined by the stimulus pair presented. Choosing stimulus A leads to correct (positive) feedback in 80% of AB trials and incorrect (negative) feedback in 20% of AB trials. The reliability of the “correct” stimulus lowers to 70% for stimulus C in CD pairs, and 60% for stimulus E in EF pairs. Over the course of training participants learn to choose stimuli A, C and E more often than B, D, or F.

Performance is evaluated after each practice block of 60 trials to ensure equivalent learning of the pairs. A criterion must be met before the test trials are administered. A different criterion for each stimulus pair was used (65% A in AB, 60% C in CD, 50% E in EF). After reaching criteria on the practice trials, participants proceed to the test session. In these trials, participants respond without feedback to the same training pairs, in addition to all novel combinations of stimuli, in random sequence. The instructions are to use "gut instinct" if they are unsure how to respond. Each test pair is presented 6 times.

The probabilistic task is difficult; in the original study in PD patients and healthy older subjects, only 59 % of subjects had useable test data. The original task was simplified in order to allow for use of more test data. The modified version used more stable probabilistic pairs (see Figure 10). In the original task, stimulus A was correct 80% of the time, and incorrect 20% of the time (vice versa for stimulus B). In the modified version, stimulus A was correct 85% of the time, and incorrect 15% of the time. The same increase in probabilistic stability was applied to the CD pair (75% and 25%) and EF pair (65% and 35%). Parallel to the increase in stability was an increase in the practice criterion. As the pairs were more stable, a higher criterion for preference (5% higher) was applied before the participants could progress to the test phase. However, as ability to meet criterion was found to be limited, test phase data was used in the analysis not only for those passing criteria for all three pairs, but was also used for those passing criteria for the 2 most stable pairs. Different Hiragana characters were used (see Figure 10) due to previous exposure of one patient to the original stimuli. Versions differing in stability were piloted on young normal participants in order to ensure that they were not performing at ceiling. The last modification was to increase the number of test trials to 90 in order to increase detection power.

Figure 10. Probability Learning Task Stimuli.

Low-verbalizable Japanese Hiragana characters were learned in three pairs. Each pair used a different probabilistic stability, where the AB pair was the most stable (A correct 85%, incorrect 15%), the CD pair was less stable (80%/20%), and the EF pair was the least stable (75%25%).



In the task, learning to choose A over B during training can result from either learning that choosing A leads to positive feedback, or that choosing B leads to negative feedback (or both). In order to determine how much learning resulted from positive versus negative feedback, the test trials include novel combinations of A or B with the stimuli from other pairs (C,D,E,F). The dependent measure is the proportion of the 90 test trials in which participants choose the good stimulus or avoid the bad stimulus. Learning from positive versus negative feedback contrasts the choice of A (the good stimulus) in novel pairings (AC, AD, AE, AF) versus the choice of B (the bad stimulus) in novel pairings (BC, BD, BE, BF).

C. RESULTS

Performance on the probability learning task (as measured by practice trial blocks to criterion on all subjects completing the task) revealed that there was a significant main effect of group [$F(2,41)=4.38$, $p < .02$]; VMPFC patients required more training trials to reach criterion (see Figure 11). Further analysis focused on only those participants reaching learning criterion on the pairs (n = 35; 80%). Of the participants passing criterion, one VMPFC-damaged patient was found to be an outlier (case with value between 1.5 and 3 interquartile ranges from either the 75th quartile or the 25th quartile of the group and more than 1.5 standard deviations from the mean group score), and was excluded from the analysis. A two-way mixed factorial ANOVA with Group (C,nVM,VM) and Learning type (Reward learning, Punishment learning) on the test phase results showed a main effect for Learning Type [$F(1,31) = 6.08$, $p = 0.019$]. Across groups, participants were better at learning from reward than from punishment (see Figure 12). There was no significant main effect for Group. However, as predicted, there was a significant interaction between Group and Learning Type [$F(2,31)=5.51$, $p < .01$] (see Figure 13). Planned post hoc t-tests revealed impaired learning from punishment compared to reward for VMPFC-damaged patients [$t = 4.93$, $p=.004$], but not for normal controls or patient controls.

Figure 11. Practice trial performance on Probability Learning task.

Shows trial blocks to criterion (n=44) for all participants completing the task (8 blocks maximum) for each group (C = normal controls, nVM = patients with damage to frontal cortex

outside VMPFC, VM = ventromedial prefrontal patients). Error bars show 95% confidence intervals.

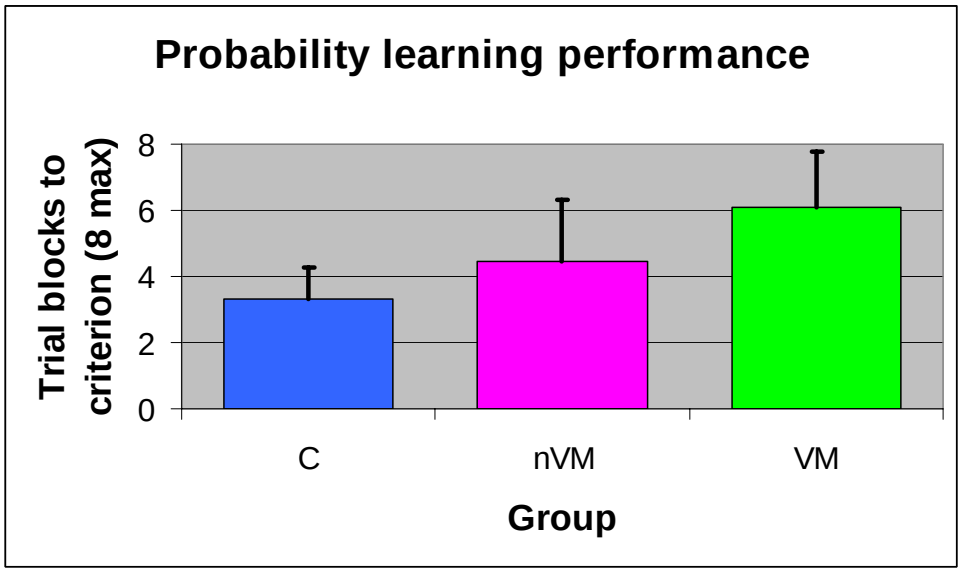


Figure 12. Probability Learning Task: Main effect for Learning Type.

Learning from reward slightly better than learning from punishment across groups. Error bars show 95% confidence intervals.

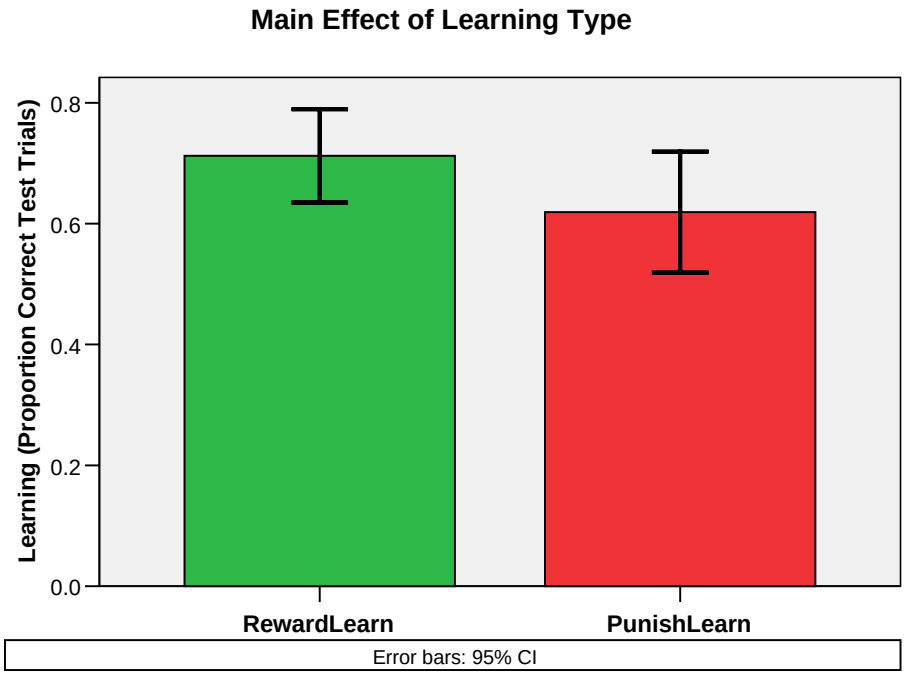
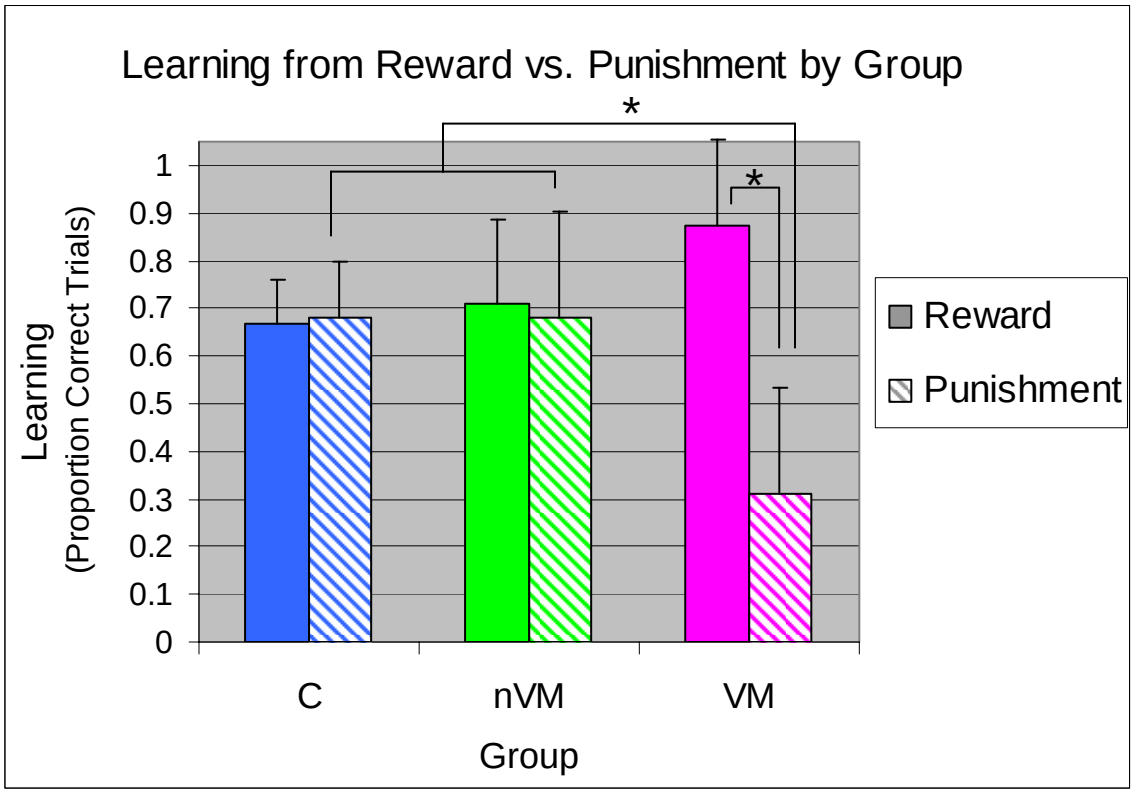


Figure 13. Interaction between Group (C,nVM,VM) and Learning Type (Reward Learning, Punishment Learning).

A significant interaction was found between Group (C = normal controls, nVM = patients with damage to frontal cortex outside VMPFC, VM = ventromedial prefrontal patients) and Learning Type (Reward Learning, Punishment Learning). VMs learned less from punishment that from reward, whereas the other groups learned equally from both forms of feedback.. Error bars show 95% confidence intervals.



D. DISCUSSION

In this study, we have shown that damage to the VMPFC results in an imbalance in the ability to learn from feedback during a probabilistic learning task. Specifically, VMPFC patients were impaired at learning from punishment compared to controls, but were not impaired at learning from reward. This impairment may also have led to overall impairment in probability learning observed in these patients, who took on average approximately 40 more trials to reach

learning criteria. An impairment in learning from negative feedback is relevant to understanding the role of VMPFC in reversal learning. Reversal learning tasks typically require learning to approach a previously punished stimulus (learn from rewards), as well as learning to reject a previously rewarded stimulus (learn from punishments). In fact, if punishment learning is completely disrupted, the participant will never be exposed to the new reward stimulus, and so will not have the ability to learn from reward. This will result in a major impairment in reversal learning. An inability to reject a previously "good" stimulus will result in perseverating on the old stimulus and impairment in reversal learning as seen by Fellows and Farah (Fellows & Farah, 2003b, 2005) in VMPFC patients.

A deficit in learning from punishment is also consistent with an account of reversal impairments in which VMPFC patients are unable to use the current negative information to reverse their previously learned stimulus-reinforcement association. Such is the case in the Iowa Gambling Task (Bechara et al., 1994), when a deck of cards has given high rewards, and then suddenly gives a very high punishment. In this case, VMPFC patients are impaired at switching from that deck to one that has given lower rewards (but also lower punishment). A specific deficit in changing response to a punishing stimulus helps explain this complex task impairment. Although a hyposensitivity to punishment has been previously tested within the gambling task framework (Bechara et al., 2000), the method used was to simply increase the amount or frequency of punishment. Merely increasing the punishment, however, may not have an effect if the problem is due to an inability to learn from such punishment. So although there was still impaired VMPFC patient performance on the variant task as the original gamble (IGT) (though performance was slightly more normal), the results do not strongly argue against a punishment-specific deficit.

There may seem to be conflict between the results of the study and the reversal literature in monkeys. If punishment learning is the specific deficit in VMPFC damage, why are lesioned monkeys showing reversal impairments when only rewards and non-rewards are used? There are several possible explanations. First of all, the part of the OFC most involved in reward reversal in monkeys was the lateral OFC (Butter, 1969b), different from the medial damage in the current study. Second, in the monkey studies, the monkeys are either receiving food or juice as a reward, or the absence of such food or juice as a non-reward. It is possible that the absence of such a salient reward acts as a punishment when compared to the alternative reward, requiring VMPFC involvement in reversal through punishment processing.

Support for a VMPFC role in punishment processing comes from both animal and human literature. In rats, recordings in VMPFC (infralimbic and prelimbic medial PFC) reveal its involvement in trace fear conditioning (Gilmartin & McEchron, 2005). In monkeys, lesions to the OFC result in blunted responses to a fear stimulus (fake snake) (Izquierdo, Suda, & Murray, 2005). In human neuroimaging, fear conditioning has been linked to the VMPFC (Tabbert, Stark, Kirsch, & Vaitl, 2005). In addition, the VMPFC has been specifically implicated in fear reversal in humans (Morris & Dolan, 2004) and fear extinction in humans (Phelps, Delgado, Nearing, & LeDoux, 2004) and rats (Morgan & LeDoux, 1995; Morgan, Schulkin, & LeDoux, 2003). In extinction, the VMPFC may be necessary during the fear acquisition stage in order for normal extinction to occur later (Morgan et al., 2003). Human imaging has also shown support for the role of VMPFC in negative feedback not involving fear. In a classification learning task using fMRI, response to negative over positive feedback activated VMPFC (Aron et al., 2004).

Not only is the VMPFC itself implicated in punishment learning, but it is linked reciprocally to the amygdala (Ghashghaei & Barbas, 2002), which is itself important in fear

conditioning (Fanselow & Poulos, 2005; LeDoux, 2003). Punishment information is conveyed through connections between the amygdale and VMPFC (Deco & Rolls, 2005; Price, Carmichael, & Drevets, 1996).

There is a possible link in punishment learning between the VMPFC and dopamine. The same task used in this study has also been used to examine the effects of dopamine on reward and punishment learning in patients with Parkinson's Disease (PD). Diminished dopamine levels are associated with relatively better learning from punishment than reward, but when dopamine levels increase following L-DOPA administration, patients learn more from reward than punishment (Frank et al., 2004). In the present experiment, damage to the VMPFC resulted in performance similar to that of Parkinson's patients on dopamine medication. Why would damage to the VMPFC be similar to an infusion of dopamine? In rats, dopamine transmission within VMPFC is implicated in both fear conditioning and extinction (Pezze & Feldon, 2004). The link of VMPFC and dopamine may be explained by results suggesting that dopamine can inhibit neural activity in medial PFC in the rat (Ferron, Thierry, Le Douarin, & Glowinski, 1984; Mantz, Milla, Glowinski, & Thierry, 1988). Therefore, if an infusion of dopamine inhibits medial PFC, it may have an impact on performance similar to that seen when the VMPFC is damaged. Further links between VMPFC and dopamine come from a neuroimaging study on humans (Aron et al., 2004), in which response to negative feedback was correlated to midbrain activity hypothesized as dopaminergic input. However, it is also possible that the similar imbalance in learning from reinforcer types is due to different underlying systems, and that there is no dopamine link in VMPFC punishment learning.

In conclusion, damage to the VMPFC was seen to impair learning from negative feedback, while learning from positive feedback was relatively intact. This suggests that the

VMPFC plays a specific role for learning from negative feedback, particularly when a change in response is needed. This role is supported by previously seen involvement of the VMPFC in reinforcer reversal learning tasks (Bechara et al., 2000; Fellows & Farah, 2003b) as well as punishment learning tasks (Aron et al., 2004; Morris & Dolan, 2004). There may also be a link between dopamine activation and VMPFC damage which results in the same relative impairment (Aron et al., 2004; Ferron et al., 1984; Mantz et al., 1988).

V. GENERAL DISCUSSION

To conclude, the VMPFC is a structure with multi-modal integration at the intersection of emotion and memory-associated regions. Damage to this region often results in subtle deficits in decision making and day-to-day functioning, with retention of normal intelligence. Several hypotheses of VMPFC function include social knowledge representation and flexible stimulus-reinforcer processing. Three experiments contrasted and examined hypotheses of the function of the VMPFC through behavioral testing of patients with VMPFC damage, patients with damage outside of VMPFC, and normal controls.

The first experiment tested the hypothesis that the VMPFC is involved in storing social knowledge (Milne & Grafman, 2001). A previous study showing lower association scores on a social category task (Milne & Grafman, 2001) was extended to include a nonsocial control task, and revealed that the difference seen in the social task was not specific to social knowledge. The evidence does support the social knowledge theory insofar as the results from the original social study were replicated. However, the additional difference in performance on the nonsocial version suggests that this hypothesis does not uniquely explain VMPFC function. So, although some evidence suggests that the VMPFC is important for social processing, that involvement

may be due to underlying involvement of the VMPFC in more simplistic processes. Comparison of negative versus positive exemplar processing did not support an impairment in processing negatively versus positively valenced stimuli. It is more likely, from the results of the other experiments, that the impairment in the IAT may reflect higher-level processing of valenced stimulus, such as that of flexibly associating stimuli and reinforcers. Such may be the case in the IAT, as the valenced attribute categories are flexibly associated with the matching (congruent blocks) versus non-matching (incongruent blocks) group categories. In fact, the performance of the VMPFC patients on the IAT was mimicked by their deficits in flexible stimulus-reinforcer processing (Expt. 3; see scatterplot in Appendix B).

The second and third experiments examined the hypothesis that the VMPFC is involved in rapid reversal of stimulus-reinforcer associations (Rolls, 2004). Both tasks were aimed at dissociating reward processing from punishment processing. The first task examined reward vs. punishment processing in a reversal gamble setting. A reward-only reversal gamble tested the specificity of the VMPFC reversal deficit to punishment situations. The normal performance observed in VMPFC patients implicates the VMPFC specifically in punishment reversals. This is opposed to their performance in gambles without reversal or without punishment. However, as all three groups were impaired, the normal performance of the VMPFC patients could be due to a floor effect, in that all three groups are close to chance. Therefore, these results do not conclusively support any theory of VMPFC function.

The second test examining rapid stimulus-reinforcer reversal tested the difference between reward versus punishment processing outside of a reversal situation. A probabilistic task (Frank et al., 2004) tested learning from punishment versus learning from reward. VMPFC

patients were impaired overall at probabilistic learning with feedback. Of those learning the task to criterion, impairment was selective for learning from punishment.

How can these results be integrated? First of all, it is clear that the VMPFC is not selective for social knowledge processing. The VMPFC is, however, implicated in flexible stimulus-reinforcer processing, and is specifically shown to be important for changing stimulus choice in response to punishment. This deficit in learning from punishment helps define patient problems seen in reversal learning and gambling tasks and is supported by neuroimaging studies on reinforcers and fear processing (Morris & Dolan, 2004; Tabbert et al., 2005). Specifically, a deficit in learning to switch choice based on punishment as seen in the Experiment 3 results will lead to an impairment in switching choice in a reversal learning task. The punishment learning deficit will result in a failure to stop responding to the old target (perseveration) when it is punished following a switch.

One question that arises from these experiments, is whether the VMPFC is solely necessary for punishment processing, but not for reward processing. Although a plentiful literature exists on the response of OFC neurons to reward and reward properties (Rolls, 2000; Rolls et al., 1996), those results do not indicate that the VMPFC is *necessary* for reward processing. Although the VMPFC may be active during reward processing, it may be necessary preferentially for punishment processing. However, as humans, our responses to reward and punishment are different depending on the context of the reinforcement (Mellers, Schwartz, Ho, & Ritov, 1997). The definition of punishment can be seen in two different ways. First, punishment can be the absolute value of a stimulus, regardless of context. For example, losing money is punishment, winning money is reward. On the other hand, punishment can be relative depending on the context. For example, winning $5 may be punishing if the alternative is to win

$150. Which way does the VMPFC code punishment and reward (or does it code reinforcers in both ways)? There is evidence from neuroimaging studies that the VMPFC codes reinforcers at least in relative terms (Breiter, Aharon, Kahneman, Dale, & Shizgal, 2001; Coricelli et al., 2005; Elliott, Friston et al., 2000). For example, a neuroimaging task involved three spinners, each with three possible monetary outcomes: one (punishment) spinner with high punishment, low punishment or no punishment, one (neutral) spinner with low punishment, low punishment, or no reinforcement, and one (reward) spinner with high reward, low reward and no reward (Breiter et al., 2001). VMPFC was mostly responsive to the outcomes in which a reward was possible (reward spinner, neutral spinner), but the worst outcome was received. This is highly similar to the third experiment (probabilistic learning), in that the VMPFC is involved when a reward is possible, but a punishment (in the spinner case, a relative punishment) is received. A different spinner task also revealed VMPFC activity to reinforcers as relative to possible outcomes (Coricelli et al., 2005). This task specifically tested regret or counterfactual thinking, in which thoughts of the alternative outcome modify reactions to the actual outcome (Mellers et al., 1997). In this task, sometimes two spinners were available. The outcome for both the chosen and alternative spinner was shown. Regret occurred when the alternative outcome was better than the chosen spinner's outcome, and relief when the alternative outcome was worse than the chosen spinner's outcome. VMPFC was parametrically more active for regret (relative punishment) than for relief. In addition, after several regret episodes, the VMPFC was active at the time of choice, indicating a possible involvement in choosing based on past reinforcer experiences. These studies indicate that VMPFC is involved in referential punishment (punishment based on context). This is especially true for situations in which the choice behavior should be adapted. In fact, activation in the VMPFC does suggest that differential

activation to a punishment is higher before a behavioral switch is made (O'Doherty et al., 2003). These results, together with the current experiments, all point to the VMPFC being involved in cases when punishment indicates a behavioral switch to a different stimulus.

Further research is important to examine this punishment learning function. One interesting extension would be to test whether the VMPFC codes punishments in an absolute sense in addition to the relative sense. Another would be to examine the importance of punishment learning in disorders related to VMPFC (depression, obsessive-compulsive disorder (OCD), post-traumatic stress disorder). For example, OCD has been linked to hyperactivity in OFC (Saxena, Bota, & Brody, 2001). Is this hyperactivity linked to changes in learning from punishment? Furthermore, reduced subgenual cingulate (subregion of VMPFC) activity in depression (Drevets, Ongur, & Price, 1998) may also be linked to changes in learning from punishment. On another tangent, monkey lesion studies could help delineate exact regions within the VMPFC involved in learning from reward versus learning from punishment.

In conclusion, these results successfully compared and examined hypotheses of VMPFC function. First, the flexible stimulus-reinforcer association function of the VMPFC was shown to be more valid than the hypothesis that the VMPFC stores social knowledge. Second, two experiments examined the stimulus-reinforcer processing deficits in VMPFC patients, revealing that learning from punishment impairment may underlie the reversal learning deficits. These two experiments suggest a special role for VMPFC in punishment processing, especially when a change in stimulus choice is indicated. These results suggest exciting new ways of explaining reversal learning and decision-making performance in VMPFC patients, as well as implications for research on other disorders affecting VMPFC.

APPENDIX A

IAT Stimuli.

Social Categories and Exemplars

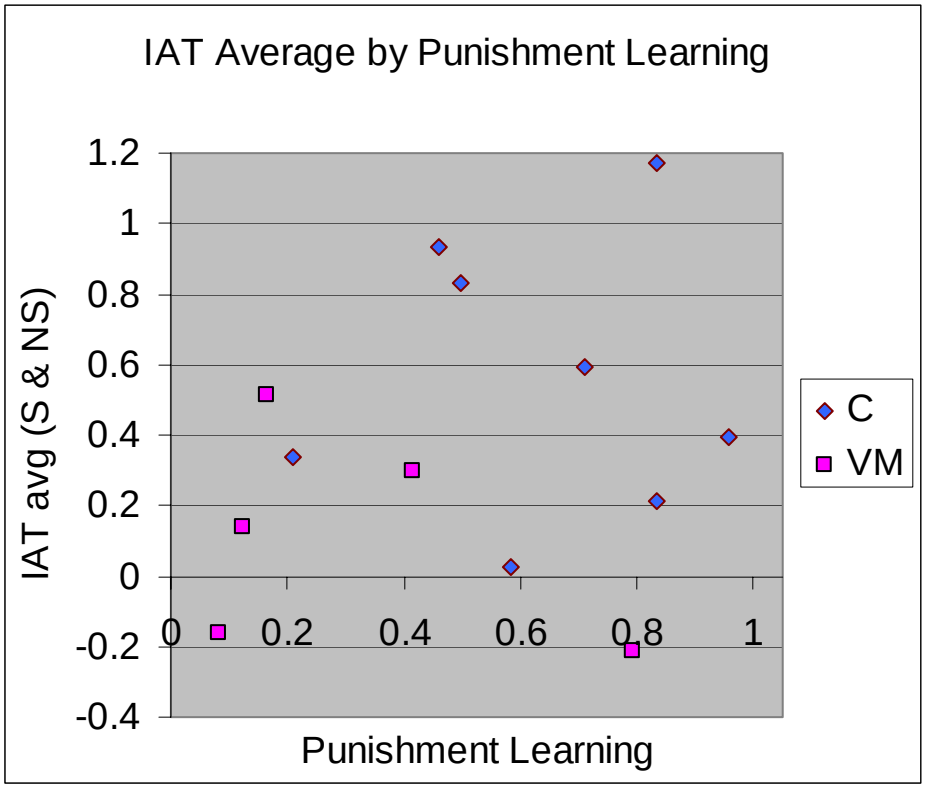
Male	Female	Strong	Weak
BRIAN	BETH	STRONG	WEAK
MATTHEW	LISA	POWER	SURRENDER
PAUL	MEG	SHOUT	TIMID
SCOTT	MARCIA	DYNAMIC	VULNERABLE
ERIC	ELAINE	WINNER	WEAKNESS
GREG	GWEN	CONFIDENT	WISPY
KEVIN	SARA	LOUD	WITHDRAW
STEVE	DIANE	BOLD	YIELD
JOHN	KAREN	SUCCEED	FAILURE
MARK	LAUREL	TRIUMPH	SHY
JASON	EVA	LEADER	FOLLOW
PETER	SANDRA	DOMINANT	LOSE
ALAN	ANN	POTENT	FRAGILE
ROBERT	SUSAN	COMMAND	AFRAID
DANIEL	KATE	ASSERT	LOSER

Nonsocial Categories and Exemplars

Flower	Insect	Pleasant	Unpleasant
CLOVER	ANT	CARESS	ABUSE
CROCUS	BEE	CHEER	ACCIDENT
DAISY	CENTIPEDE	DIAMOND	ASSAULT
GLADIOLA	COCKROACH	DIPLOMA	BOMB
HYACINTH	CRICKET	GENTLE	CRASH
IRIS	FLEA	GIFT	DISASTER
LILAC	FLY	HEAVEN	DIVORCE
MAGNOLIA	GNAT	LAUGHTER	FILTH
PANSY	HORNET	LOYAL	JAIL
PEONY	HORSEFLY	LUCKY	POISON
PETUNIA	LOCUST	MIRACLE	POLLUTE
POPPY	MAGGOT	PARADISE	POVERTY
ROSE	SPIDER	RAINBOW	ROTTEN
VIOLET	TARANTULA	SUNRISE	SICKNESS
ZINNIA	WEEVIL	VACATION	STINK

APPENDIX B

Scatterplot: IAT Effect by Punishment Learning



BIBLIOGRAPHY

Anderson, A. K., Christoff, K., Stappen, I., Panitz, D., & Ghahremani, D. G. (2003). Dissociated neural representations of intensity and valence in human. *Nat Neurosci, 6*(2), 196-202.

Anderson, J. R., de Monte, M., & Kempf, J. (1996). Discrimination learning and multiple reversals in young adult and older monkeys (*Macaca arctoides). Quarterly Journal of Experimental Psychology, 49B*(3), 193-200.

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *J Neurophysiol, 92*(2), 1144-1152.

Barrash, J., Tranel, D., & Anderson, S. W. (2000). Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. *Dev Neuropsychol, 18*(3), 355-381.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience, 18*(1), 428-437.

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain: A Journal of Neurology, 123*(11), 2189-2202.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*(6), 1142-1163.

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., et al. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology, 12*(3), 446-458.

Breiter, H. C., Aharon, I., Kahneman, D., Dale, A., & Shizgal, P. (2001). Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron, 30*(2), 619-639.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: JA Barth.

Buckner, R. L., Raichle, M. E., Miezin, F. M., & Petersen, S. E. (1996). Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J Neurosci, 16*(19), 6219-6235.

Butter, C. M. (1969a). Impairments in selective attention to visual stimuli in monkeys with inferotemporal and lateral striate lesions. *Brain Res, 12*(2), 374-383.

Butter, C. M. (1969b). Perseveration in extinction and in discrimination reversal tasks following selective frontal ablations in *Macaca Mulatta*. *Physiology and Behavior, 4*, 163-171.

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J Comp Neurol, 363*(4), 615-641.

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *J Comp Neurol, 371*(2), 179-207.
Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., et al. (2000). Parsing executive processes: strategic vs. evaluative functions of the anterior cingulate cortex. *Proc Natl Acad Sci U S A, 97*(4), 1944-1948.
Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: a neuroimaging study of choice behavior. *Nat Neurosci, 8*(9), 1255-1262.
Critchley, H. D., & Rolls, E. T. (1996). Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J Neurophysiol, 75*(4), 1673-1686.
Damasio, A. R. (1994). Descartes' error and the future of human life. *Sci Am, 271*(4), 144.
Damasio, A. R., & Van Hoesen, G. W. (1983). Emotional disturbance associated with focal lesions of the limbic frontal lobe. In K. Heilman & P. Satz (Eds.), *Neuropsychology of human emotion*. New York: Guilford Press.
Davidson, R. J. (1998). Anterior electrophysiological asymmetries, emotion, and depression: conceptual and methodological conundrums. *Psychophysiology, 35*(5), 607-614.
Davidson, R. J., Jackson, D. C., & Kalin, N. H. (2000). Emotion, plasticity, context, and regulation: perspectives from affective neuroscience. *Psychol Bull, 126*(6), 890-909.
Davidson, R. J., & Sutton, S. K. (1995). Affective neuroscience: the emergence of a discipline. *Curr Opin Neurobiol, 5*(2), 217-224.
Decety, J., & Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia, 41*(2), 127-138.
Deco, G., & Rolls, E. T. (2005). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cereb Cortex, 15*(1), 15-30.
Devine, P. G. (Ed.). (1989). *Automatic and controlled processes in prejudice: The role of stereotypes and personal beliefs*: Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, England.
Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature, 380*(6569), 69-72.
Drevets, W. C., Ongur, D., & Price, J. L. (1998). Reduced glucose metabolism in the subgenual prefrontal cortex in unipolar depression. *Mol Psychiatry, 3*(3), 190-191.
Elliott, R., Dolan, R. J., & Frith, C. D. (2000). Dissociable functions in the medial and lateral orbitofrontal cortex: Evidence from human neuroimaging studies. *Cerebral Cortex, 10*(3), 308-317.
Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience, 20*(16), 6159-6165.
Fanselow, M. S., & Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annu Rev Psychol, 56*, 207-234.
Fellows, L. K., & Farah, M. J. (2003a). *Affective shifting in the Iowa gambling task: Gambling performance in humans with ventromedial frontal lobe damage reflects a fundamental reversal learning deficit.* Paper presented at the Society for Neuroscience meeting, New Orleans.
Fellows, L. K., & Farah, M. J. (2003b). Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. *Brain, 126*(Pt 8), 1830-1837.

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cereb Cortex, 15*(1), 58-63.

Ferron, A., Thierry, A. M., Le Douarin, C., & Glowinski, J. (1984). Inhibitory influence of the mesocortical dopaminergic system on spontaneous activity or excitatory response induced from the thalamic mediodorsal nucleus in the rat medial prefrontal cortex. *Brain Res, 302*(2), 257-265.

Fiez, J. A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping, 5*(2), 79-83.

Folstein, M. F., Robins, L. N., & Helzer, J. E. (1983). The Mini-Mental State Examination. *Arch Gen Psychiatry, 40*(7), 812.

Frank, M. J., Seeberger, L. C., & O'Reilly R, C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science, 306*(5703), 1940-1943.

Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proc Natl Acad Sci U S A, 95*(3), 906-913.

Gaffan, D., Murray, E. A., & Fabre-Thorpe, M. (1993). Interaction of the amygdala with the frontal lobe in reward memory. *Eur J Neurosci, 5*(7), 968-975.

Ghashghaei, H. T., & Barbas, H. (2002). Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience, 115*(4), 1261-1279.

Gilmartin, M. R., & McEchron, M. D. (2005). Single neurons in the medial prefrontal cortex of the rat exhibit tonic and phasic coding during trace fear conditioning. *Behav Neurosci, 119*(6), 1496-1510.

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology, 70*(3), 491-512.

Goldapple, K., Segal, Z., Garson, C., Lau, M., Bieling, P., Kennedy, S., et al. (2004). Modulation of cortical-limbic pathways in major depression: treatment-specific effects of cognitive behavior therapy. *Arch Gen Psychiatry, 61*(1), 34-41.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol, 74*(6), 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *J Pers Soc Psychol, 85*(2), 197-216.

Henriques, J. B., & Davidson, R. J. (1991). Left frontal hypoactivation in depression. *J Abnorm Psychol, 100*(4), 535-545.

Hikosaka, K., & Watanabe, M. (2000). Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb Cortex, 10*(3), 263-271.

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., et al. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain, 126*(Pt 7), 1691-1712.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *J Cogn Neurosci, 16*(3), 463-478.

Iversen, S. D., & Mishkin, M. (1970). Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. *Exp Brain Res, 11*(4), 376-386.

Izquierdo, A., & Murray, E. A. (2005). Opposing effects of amygdala and orbital prefrontal cortex lesions on the extinction of instrumental responding in macaque monkeys. *Eur J Neurosci, 22*(9), 2341-2346.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2004). Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J Neurosci, 24*(34), 7540-7548.

Izquierdo, A., Suda, R. K., & Murray, E. A. (2005). Comparison of the effects of bilateral orbital prefrontal cortex lesions and amygdala lesions on emotional responses in rhesus monkeys. *J Neurosci, 25*(37), 8534-8542.

Jones, B., & Mishkin, M. (1972). Limbic lesions and the problem of stimulus--reinforcement associations. *Exp Neurol, 36*(2), 362-377.

Kahneman, D., & Tversky, A. (1991). Loss aversion in riskless choice: a reference-dependent model. *Q. J. Econ., 106*, 1039-1061.

LeDoux, J. (2003). The emotional brain, fear, and the amygdala. *Cell Mol Neurobiol, 23*(4-5), 727-738.

Ledoux, J. E., & Muller, J. (1997). Emotional memory and psychopathology. *Philos Trans R Soc Lond B Biol Sci, 352*(1362), 1719-1726.

Mantz, J., Milla, C., Glowinski, J., & Thierry, A. M. (1988). Differential effects of ascending neurons containing dopamine and noradrenaline in the control of spontaneous activity and of evoked responses in the rat prefrontal cortex. *Neuroscience, 27*(2), 517-526.

Mayberg, H., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., et al. (1999). Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. *Am J Psychiatry, 156*(5), 675-682.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science, 8*(6), 423-429.

Milne, E., & Grafman, J. (2001). Ventromedial prefrontal cortex lesions in humans eliminate implicit gender stereotyping. *J Neurosci, 21*(12), RC150.

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. *Journal of Cognitive Neuroscience, 9*(6), 788-798.

Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci, 22*(7), 2730-2736.

Morgan, M. A., & LeDoux, J. E. (1995). Differential contribution of dorsal and ventral medial prefrontal cortex to the acquisition and extinction of conditioned fear in rats. *Behav Neurosci, 109*(4), 681-688.

Morgan, M. A., Schulkin, J., & LeDoux, J. E. (2003). Ventral medial prefrontal cortex and emotional perseveration: the memory for prior extinction training. *Behav Brain Res, 146*(1-2), 121-130.

Morris, J. S., & Dolan, R. J. (2004). Dissociable amygdala and orbitofrontal responses during reversal fear conditioning. *Neuroimage, 22*(1), 372-380.

Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: a meta-analysis. *Cogn Affect Behav Neurosci, 3*(3), 207-233.

Nieuwenhuis, S., Slagter, H. A., von Geusau, N. J., Heslenfeld, D. J., & Holroyd, C. B. (2005). Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. *Eur J Neurosci, 21*(11), 3161-3168.

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *Journal of Neuroscience, 23*(21), 7931-7939.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci, 4*(1), 95-102.

Ongur, D., Drevets, W. C., & Price, J. L. (1998). Glial reduction in the subgenual prefrontal cortex in mood disorders. *Proc Natl Acad Sci U S A, 95*(22), 13290-13295.

Ongur, D., Ferry, A. T., & Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *J Comp Neurol, 460*(3), 425-449.

Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb Cortex, 10*(3), 206-219.

Pears, A., Parkinson, J. A., Hopewell, L., Everitt, B. J., & Roberts, A. C. (2003). Lesions of the orbitofrontal but not medial prefrontal cortex disrupt conditioned reinforcement in primates. *J Neurosci, 23*(35), 11189-11201.

Pezze, M. A., & Feldon, J. (2004). Mesolimbic dopaminergic pathways in fear conditioning. *Prog Neurobiol, 74*(5), 301-320.

Phelps, E. A., Delgado, M. R., Nearing, K. I., & LeDoux, J. E. (2004). Extinction learning in humans: role of the amygdala and vmPFC. *Neuron, 43*(6), 897-905.

Price, J. L., Carmichael, S. T., & Drevets, W. C. (1996). Networks related to the orbital and medial prefrontal cortex; a substrate for emotional behavior? *Prog Brain Res, 107*, 523-536.

Rauch, S. L., Savage, C. R., Alpert, N. M., Fischman, A. J., & Jenike, M. A. (1997). The functional neuroanatomy of anxiety: A study of three disorders using positron emission tomography and symptom provocation. *Biological Psychiatry, 42*(6), 446-452.

Remijnse, P. L., Nielen, M. M., Uylings, H. B., & Veltman, D. J. (2005). Neural correlates of a reversal learning task with an affectively neutral baseline: an event-related fMRI study. *Neuroimage, 26*(2), 609-618.

Roesch, M. R., & Olson, C. R. (2004). Neuronal activity related to reward value and motivation in primate frontal cortex. *Science, 304*(5668), 307-310.

Rolls, E. T. (1996). The orbitofrontal cortex. *Philos Trans R Soc Lond B Biol Sci, 351*(1346), 1433-1443; discussion 1443-1434.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cereb Cortex, 10*(3), 284-294.

Rolls, E. T. (2004). The functions of the orbitofrontal cortex. *Brain Cogn, 55*(1), 11-29.

Rolls, E. T., Critchley, H. D., Browning, A. S., Hernadi, I., & Lenard, L. (1999). Responses to the sensory properties of fat of neurons in the primate orbitofrontal cortex. *J Neurosci, 19*(4), 1532-1540.

Rolls, E. T., Critchley, H. D., Mason, R., & Wakeman, E. A. (1996). Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J Neurophysiol, 75*(5), 1970-1981.

Rorden, C., & Brett, M. (2000). Stereotaxic display of brain lesions. *Behavioural Neurology, 12*(4), 191-200.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26*(11), 1315-1328.
Saxena, S., Bota, R. G., & Brody, A. L. (2001). Brain-behavior relationships in obsessive-compulsive disorder. *Semin Clin Neuropsychiatry, 6*(2), 82-101.
Schoenbaum, G., Chiba, A. A., & Gallagher, M. (1998). Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nature Neuroscience, 1*(2), 155-159.
Schultz, W., Tremblay, L., & Hollerman, J. R. (1998). Reward prediction in primate basal ganglia and frontal cortex. *Neuropharmacology, 37*(4-5), 421-429.
Schultz, W., Tremblay, L., & Hollerman, J. R. (2000). Reward processing in primate orbitofrontal cortex and basal ganglia. *Cereb Cortex, 10*(3), 272-284.
Tabbert, K., Stark, R., Kirsch, P., & Vaitl, D. (2005). Hemodynamic responses of the amygdala, the orbitofrontal cortex and the visual cortex during a fear conditioning paradigm. *Int J Psychophysiol, 57*(1), 15-23.
Thorpe, S. J., Rolls, E. T., & Maddison, S. (1983). The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res, 49*(1), 93-115.
Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature, 398*(6729), 704-708.
van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage, 14*(6), 1302-1308.
Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31*(2), 329-338.
Walker, A. E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. *J Comp Neurol, 73*, 53-86.
Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature, 382*(6592), 629-632.